The Thesis Committee for Zeina Wafa
Certifies that this is the approved version of the following thesis:

A Latent-Segmentation Based Approach to Investigating the Spatial Transferability of Activity-Travel Models

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Chandra Bhat

____
Stephen Boyles

A Latent-Segmentation Based Approach to Investigating the Spatial Transferability of Activity-Travel Models

by

Zeina Wafa, B.S.E.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

December 2014

Dedication

To my parents, sisters, and Rami.

Acknowledgements

I am most grateful to my friend Sebastian Astroza Tagle for supporting me until the finish line. Without his help, the modeling work involved in this thesis would not have been successfully completed. His selfless efforts towards making my life a lot easier will remain dear to my heart forever.

Special thanks to Sergio Martinez for being my guide throughout graduate school, Carine Choubassi for being my closest friend in Austin, Mayra Martinez for reminding me to always take it easy, Stasa Zivojnovic for all the laughs, Rita Nehme for the perpetual inspiration, Weam Alawar for much needed wisdom, Ignacio de Sande Palma for being a friend in need, Sabine Francis for being my go-to-gal, Maher Saeed for putting up with all the nagging from abroad and Ahmad Al-Rushaidan for infinity.

I would also like to thank my parents for their unconditional love and support and my sisters, Hala and Jana, whom I hope I will be acknowledged in their theses and dissertations one day.

Finally, I would like to extend my heartfelt gratitude to my advisor, Dr. Chandra Bhat, for his inspiration to always seek knowledge and indulge in research. Special thanks to Lisa Macias for being my resourceful reference. Finally, I would like to thank Dr. Stephen Boyles for taking the time to read this thesis and provide valuable comments.

Hook'em!

Abstract

A Latent-Segmentation Based Approach to Investigating the Spatial Transferability of Activity-Travel Models

Zeina Wafa, M.S.E.
The University of Texas at Austin, 2014

Supervisor: Chandra Bhat

Spatial transferability of travel demand models has been an issue of considerable interest, particularly for small and medium sized planning areas that often do not have the resources and staff time to collect large scale travel survey data and estimate model components native to the region. With the advent of more sophisticated microsimulation-based activity-travel demand models, the interest in spatial transferability has surged in the recent past as smaller metropolitan planning organizations seek to take advantage of emerging modeling methods within the limited resources they can marshal. Traditional approaches to identifying geographical contexts that may borrow and transfer models between one another involve the exogenous *a priori* identification of a set of variables that are used to characterize the similarity between geographic regions. However, this *ad hoc* procedure presents considerable challenges as it is difficult to identify the most appropriate criteria a priori. To address this issue, this thesis proposes a latent segmentation approach whereby the most appropriate criteria for identifying areas with similar profiles are determined endogenously within the model estimation phase, customized for every model type. The end products are a set of optimal similarity measures that link regions to one another as well as a fully transferred model, segmented to account for heterogeneity in the population. The methodology is demonstrated and its efficacy established through a case study that utilizes the National Household Travel Survey (NHTS) dataset for information on weekday activities unemployed individuals within 9 regions in the states of California and Florida engage in. A multiple discrete continuous extreme value (MDCEV) model is developed that simulates the discrete nature of activity selection as well as the continuous nature of activity participation. The estimated model is then applied onto the Austin–San Marcos MSA, a context withheld from the original estimation in order to assess its performance. The performance of the segmented model was then examined vis-à-vis that of other models that are similar to the local region in only one dimension. It is found that the methodology offers a robust mechanism for identifying latent segments and establishing criteria for transferring models between areas.

Table of Contents

List of Tables x

List of Figures xi

Chapter 1: Introduction 1

Chapter 2: Data Description..8

Chapter 3: Modeling Methodology...13

3.1. Multiple Discrete-Continuous Extreme Value Model 13

3.2. Segment-Specific Model Formulation 15

3.3. Segment Assignment Formulation...19

3.4. Measures of Goodness-of-Fit...20

Chapter 4: Model Estimation Results 22

4.1. Latent Segmentation Results..22

4.2. Comparison of Endogenous and Exogenous Segmentation Schemes ...28

4.3. Spatial Transferability Based on Latent Segmentation.........................32

Chapter 5: Summary and Conclusions..41

Appendix..44

Bibliography 65

List of Tables

Table 1: Descriptive Statistics of the Sample by State 11

Table 2: The Latent Segmentation Model and Characterization of Three Segments 26

Table 3: Comparison of the Endogenous Segmentation Model with One-Way and Two-Way Exogenous Segmentation Models 30

Table 4: Comparison of the Predictions by the MDCEV Model for Austin–San Marcos and the Transferred Segmented Model 34

Table 5: Goodness-of-fit Measures for the Locally Estimated and Transferred Models 37

Table 6: Segment 1 Model Estimation Results 44

Table 7: Segment 2 Model Estimation Results 47

Table 8: Segment 3 Model Estimation Results 50

Table 9: MDCEV Model Results for Austin–San Marcos 53

List of Figures

Figure 1: Comparison of Predictions of In-Home Activities by the Estimated and Transferred Model against the Observed In-Home Activity Participation 56

Figure 2: Comparison of Predictions of Out-of-Home Shopping Activities by the Estimated and Transferred Model against the Observed Out-of-Home Shopping Activity Participation 57

Figure 3: Comparison of Predictions of Out-of-Home Maintenance Activities by the Estimated and Transferred Model against the Observed Out-of-Home Maintenance Activity Participation 58

Figure 4: Comparison of Predictions of Out-of-Home Social/Recreational Activities by the Estimated and Transferred Model against the Observed Out-of-Home Social/Recreational Activity Participation 59

Figure 5: Comparison of Predictions of Out-of-Home Active Recreation Activities by the Estimated and Transferred Model against the Observed Out-of-Home Active Recreation Activity Participation... 60

Figure 6: Comparison of Predictions of Out-of-Home Medical Activities by the Estimated and Transferred Model against the Observed Out-of-Home Medical Activity Participation 61

Figure 7: Comparison of Predictions of Out-of-Home Eating Activities by the Estimated and Transferred Model against the Observed Out-of-Home Eating Activity Participation 62

Figure 8: Comparison of Predictions of Out-of-Home Pickup/Drop-off Activities by the Estimated and Transferred Model against the Observed Out-of-Home Pickup/Drop-off Activity Participation... 63

Figure 9: Comparison of Predictions of Out-of-Home Other Activities by the Estimated and Transferred Model against the Observed Out-of-Home Other Activity Participation .... .64

Chapter 1: Introduction

There is considerable interest among the transportation planning and modeling community in the notion of spatial transferability of travel demand models. Spatial transferability of a model refers to the ability to use a model that was estimated in one context in a different application context, and obtain useful results that approximate locally observed behavior in the application context. While it is generally considered good practice to develop models based on locally collected data, some regions, particularly small and medium-sized planning organizations, may not have the resources and staff time necessary to undertake large scale survey data collection efforts and thus borrow models from other regions (Sikder et al., 2012). When such model transfer is considered, it is important to ensure that the transferred model offers useful and valid information in the application context (Koppelman and Wilmot, 1982).

There are two main approaches adopted in the literature for spatial transferability: application-based transferability and estimation-based transferability. In the context of the former, a model is developed and estimated based on data from one region, denoted as the *estimation context*. The model is then applied in the *application context* where its predictive abilities are assessed. Models can be simply transferred as is (naïve transfer) or may be calibrated and modified to better fit the data in the application context. Updating methods include updating the constants of the utility functions, updating the scale of the random error terms as well as the constants of the utility functions, Bayesian methods, combined transfer estimation and joint context estimation. Numerous studies in the literature adopt this approach to model transferability (see Koppleman and Rpse, 1983; Koppelman and Wilmot, 1986; Koppelman and Pas, 1986; Wilmot, 1995; Arentze et al., 2002; Bekhor and Prato, 2009; Nowrouzian and Srinivasan, 2012; Sikder and Pinjari, 2013a; Sikder et al., 2013b). On the other hand, in the method of estimation-based transferability, also referred to as joint context estimation, the model is developed and estimated based on data from both contexts, the *estimation* and *application* areas. "Difference" parameters are included in the model to account for the fact that data comes from more than one source. Statistical tests are then carried out to determine whether the difference variables are significant, i.e., whether the parameters on a certain variable are essentially different between the two (or more) contexts. In this approach, common parameters can be estimated for variables for which data is limited (see Karasmaa 2001; Sikder et al., 2013b; Bowman et al., 2014). This approach also allows for statistical tests on the differences between coefficients and thus permits a wide range of hypotheses. Another advantage is that it allows for the transferability assessment of parameters associated with specific variables, while the application-based transferability approach focuses exclusively on assessing transferability of the model as a whole. One disadvantage may be that, since this approach employs statistical methods to assess transferability, transferability results may be highly sensitive to sample size or consistency in data sources (Bowman et al., 2014).

Traditionally, in the absence of data native to the region, the transfer method is based on identifying another metropolitan area that is similar to the local conditions in one or more ways. The literature identifies a number of factors that may be considered when determining the nature of contexts for which transferability can be successfully achieved. The literature suggests similar transit service quality (see McComb, 1986; Stopher et al., 2003; Mohammadian and Zhang, 2007), income levels (see Caldwell and Demetsky, 1980; Wilmot, 1996; Reuscher et al., 2002), demographic characteristics (see Caldwell and Demetsky, 1980; Mohammadian and Zhang, 2007), city socio-economic composition (Caldwell and Demetsky, 1980), and area size and type (see McComb, 1986; Stopher et al., 2003; Reuscher et al., 2002; Everett, 2009) are significant determinants of the success of spatial transferability of a model. Also, it has been found that intra-state transferability outperforms inter-state transferability (see Sikder and Pinjari, 2013a; Bowman et al., 2014). Accordingly, in these and other previous studies dealing with spatial transferability of models, similar contexts or planning areas worthy of model transfer have been defined based on a set of exogenously specified criteria. The problem with this approach is three fold: First, it defines a priori what the parameter(s) is that define(s) the measure of similarity between the local region and the transfer region. Second, it assumes that a single uniform set of parameters measuring similarity is at work regardless of the type of model being borrowed. Third, the transfer is based on centralized measures of tendency between the local region and the transfer region.

The first problem refers to the fact that the parameters of similarity are exogenously identified. However, it is likely that similarity between regions is a multi-dimensional measure. For example, it may not be adequate for a local region to borrow a model from another region that is similar only on the residential density dimension. One way to accommodate this multi-dimensional similarity within the exogenous approach is to partition regions along all potentially relevant dimensions. However, a practical problem with this "full-dimensional" exogenous transfer scheme is that there may not be a unique region that lies at the intersection of all the dimensions as the local region. To overcome this limitation, it is typical to consider only one or two dimensions that are a priori designated as the most important measures of closeness. The disadvantage is that closeness on a whole host of potentially important dimensions is discarded away. In addition, an intrinsic problem with all exogenous transfer approaches is that the threshold values of the continuous variables (for example, residential density or employment density) which define closeness or similarity have to be established in a rather *ad hoc* fashion.

The second problem is that the exogenous approach, because it is ad hoc in its identification of what constitutes similarity, uses the same set of similarity dimensions regardless of the type of model being transferred. On the other hand, it may be that residential density is a better measure of similarity when transferring a model associated with activity time-use, while the availability of specific forms of transit as in the local region may be the key similarity measure for mode choice. What would be helpful here is a way to extract information regarding similarity in another way that is customized to the model to be transferred.

The third problem is that there are likely to be different spatial pockets within metropolitan areas that are quite different from one another on the similarity measures used in exogenous schemes. For example, different areas in the same city exhibit different residential densities. However, the exogenous schemes use a single central measure to characterize entire metropolitan regions (such as a mean residential density measure), and use that central tendency to determine the region that is closest to the local region. However, the local region may have pockets that are highly dense that reveal individual and household-level activity-travel behavior patterns similar to dense pockets in other regions, while also having pockets of low density in which the activity-travel behavior patterns are similar to low density pockets in other regions. What would be nice is to allow for this heterogeneity spatial characteristics within the local region.

All three of the problems above can be resolved using a novel approach to transferability that is based on an endogenous transfer approach. The novelty is that we borrow from all the data that is available from all other regions that have information on the relevant activity-travel dimension of interest (and appropriate exogenous variables), rather than a priori decide a single region to borrow from. In this approach, there is no need to limit the dimensions of similarity to one or two, because the concept of similarity is simultaneously based on multiple dimensions. In particular, a limited number of latent segments is derived, specific to each kind of model being transferred, by characterizing each latent segment by the entire set of potentially relevant similarity variables. The appropriate number of latent segments that is appropriate for a specific activity-travel dimension of interest is determined statistically by successively adding an additional segment till a point is reached where an additional segment does not result in a significant improvement in fit. Individuals, based on their location characteristics as captured in the potentially relevant similarity variable measures, are assigned to segments in a probabilistic fashion. That is, each latent segment refers to an optimal combination of location characteristics that make individuals within that segment behave similarly on the activity-travel dimension measure of interest. The endogenous approach jointly determines the number of segments, the assignment of individuals to segments, and segment-specific choice model parameters. Since this approach identifies segments without requiring a multi-way partitioning based on all potentially relevant similarity measures as in the full-dimensional exogenous transfer method, it allows the use of all similarity variables in practice. Because the similarity-based latent segmentation scheme is estimated jointly with the main activity-travel dimension model of interest, it is immediately customized to the task at hand. Finally, by using data from a host of different regions, it captures the heterogeneity in locational characteristics and its impacts on the activity-travel behavior dimension of interest. This allows us to recognize the heterogeneity that exists within different spatial pockets of the local region.

The model considered in this study is similar to the activity generation and time-use model discussed in Sikder and Pinjari (2013a). However, rather than assessing spatial transferability via naïve transfer or transfer with constants update – as was done in their paper – this study aims to study spatial transferability in an estimation-based context where the latent classification of the dataset results in endogenously identifying rather homogeneous segments comprising different but 'similar' regions based on a number of criteria which the model estimation yields. It is noteworthy that the identified drawbacks of the estimation-based transferability approach do not pose a challenge in this research as the data is derived from the same source and the sample size is large enough.

The remainder of this thesis is organized as follows. The second section offers a description of the dataset used in this study. The modeling methodology is presented in the third section. Model estimation results are presented in the fourth section, while an assessment of the latent segments and spatial transferability is furnished in the fifth section. The sixth and final section presents conclusions.

Chapter 2: Data Description

The data used in this study is drawn from the 2009 National Household Travel Survey (NHTS). The analysis considers weekday activity participation of unemployed adults (18 years or above). In order to prepare the dataset for this study, extensive data filtering was performed. Records with incomplete or missing information, weekend activity-travel records, and long distance travel (150 miles or longer) were removed from the dataset. The out-of-home activities were classified into eight categories: shopping, maintenance, social/recreational, active recreation, medical, eat out, pickup/drop-off, and others. Similar activities were aggregated in terms of their dwell times. For example, if an individual performed a shopping activity for 30 minutes and another shopping activity for 50 minutes, the aggregation resulted in two shopping activities with 80 minutes of total shopping dwell time. The total in-home activities dwell time was inferred by subtracting the total out-of-home activities dwell time, the total travel time, and sleep time (taken to be 520 minutes according to the 2009 American Time Use Survey) from the total time of 24 hours in a day. After filtering out inconsistent records (those with dwell times and travel times adding up to more than 24 hours a day[1] and those with combinations of dwell times and travel times that lead to negative in-home activities dwell time), and removing duplicate entries for the same individual, the final dataset included records for 28,264

[1] It is noteworthy that there was no tolerance allowed for this filtering. For example, if dwell times and travel times summed up to even one minute more than 24 hours, the record was removed from the dataset.

individuals belonging to 39 different states. In the interest of computational time considerations, this thesis focuses on weekday daily activity-travel information pertaining only to the states of California and Florida with a sample size of 10,649 individuals.

Variables of interest for the model estimation are activity purpose, activity dwell time, travel time, age of respondent, sex of respondent, household income, race of household respondent, driver status of respondent, number of drivers in the household, number of household members, number of workers in the household, number of adults in the household, life cycle classification of the household, travel day of the week, highest grade attained by respondent, residential density per square mile, employment density per square mile, and the rail status in the metropolitan area where the household is located.

Table 1 presents the socio-economic and activity engagement characteristics of the survey data sample. The sample contains activity participation information from nine regions: Los Angeles – Riverside – Orange County, CA; Sacramento – Yolo, CA; San Diego, CA; San Francisco – Oakland – San Jose, CA; Jacksonville, FL; Miami – Fort Lauderdale, FL; Orlando, FL; Tampa – St Petersburg – Clearwater, FL; and West Palm Beach – Boca Raton, FL. The respective state samples are significantly different from one another. For example, the age distribution shows a higher percentage of young and middle aged people (18 – 54 years) in California than in Florida, a higher percentage of older individuals (55+ years) in Florida than in California. This is consistent with the notion that Florida is a popular destination for retirees and hence there is a relatively high proportion of older individuals. There is a higher percentage of people with a bachelor's degree or higher in California than in Florida. Individuals belonging to the California sample seem to be wealthier than those in the Florida sample (35.9 percent with income greater than $75,000 in California compared to 26.1 percent for Florida), although this should be interpreted in the context of the cost of living differential between the two states. Cost of living is generally higher in California than in Florida. These differences in socio-demographic characteristics between the two states may contribute to individuals residing in different areas exhibiting varying intrinsic preferences for activity participation and time-use. Therefore, it may be expected that models estimated on individual segments will provide more robust predictions than a model estimated on the dataset as a whole (in replicating observed activity-travel patterns in each geographical region).

Table 1: Descriptive Statistics of the Sample by State

Characteristic	**California**	**Florida**	**Total**
Sample Size	7,048	3,601	10,649
Gender: Male	40.7%	41.2%	40.9%
Age: 18 – 29 years	8.0%	4.3%	6.6%
Age: 30 – 54 years	22.7%	16.6%	20.6%
Age: 55 – 64 years	18.6%	18.2%	18.5%
Age: 65 – 74 years	25.5%	28.3%	26.5%
Age: ≥75 years	25.2%	32.6%	27.8%
Race: White	78.8%	87.9%	81.9%
Race: Black	3.8%	6.7%	4.8%
Race: Other	17.4%	5.4%	13.3%
Driver Status	91.3%	90.9%	91.2%
Education: High school level or lower	29.5%	37.5%	32.2%
Education: Some college level	32.1%	28.5%	30.9%
Education: Bachelor's level or higher	38.4%	34.0%	36.9%
Income: <25 K	18.5%	26.3%	21.1%
Income: 25 K – 50 K	27.7%	31.5%	29.0%
Income: 50 K – 75 K	17.9%	16.1%	17.3%
Income: ≥75 K	35.9%	26.1%	32.6%
Average Household Size	2.5	2.2	2.4
Average Number of Drivers	1.9	1.8	1.9
Average Number of Activities	3.0	3.0	3.0
Average Activity Duration (min)[a]			
Home	702.5	705.8	703.6
Shop	60.3	61.1	60.6
Maintenance	31.3	31.8	31.5
Social	161.9	156.7	160.1
Active	83.7	79.8	82.4
Medical	79.2	87.6	82.0
Eat-out	63/4	65.2	64.0
Pick-up	44.7	43.6	44.3
Other	148.9	121.1	139.5

[a]average durations are computed only on the portion of the sample that participated in each of the activities

The dependent variable in this modeling effort is individual-level activity generation and time-use. As mentioned previously, there are eight types of out-of-home activities. Moreover, an individual can choose the degree to which he/she participates in the chosen activity – represented by the activity dwell time (in minutes). Table 1 shows the variability in the dependent variable characteristics across the states in the dataset.

The information presented reflects the average number of activities an unemployed adult undertakes on a weekday, as well as the average duration an individual participates in a certain type of activity (by state and for the dataset as a whole). It is seen that individuals exhibit considerable similarity in their activity engagement and time use profiles, albeit with a few notable differences. For example, individuals in Florida spend more time for medical related activities (consistent with the older age profile of the survey sample), while California residents spend more time for social and other activities. Residents in California also show marginally higher levels of time use for active recreational pursuits.

Chapter 3: Modeling Methodology

This section presents an overview of the modeling methodology adopted in this study. The methodology includes segment-specific model formulation and assignment components that provide the ability to identify latent segments endogenously and then assign regions to different segments based on the endogenously identified criteria.

3.1. Multiple Discrete-Continuous Extreme Value Model

Single discrete choice models, such as multinomial logit (MNL) and multinomial probit (MNP), are typically utilized to model a decision making process where decision makers choose one alternative from a set of feasible alternatives. Some choice processes, however, involve the choice of multiple alternatives from the universal choice set of alternatives. An example of such a multiple-discrete choice process includes the choice of multiple vehicle types from an array of vehicle types available in the market (for example, a household may own both a car and a minivan) or the array of food choices that a household consumes. In addition to choosing multiple alternatives, an individual or household may consume each of the chosen alternatives to different degrees. Pairing multiple-discrete choice process with the continuous consumption component leads to the formulation of Kuhn-Tucker demand functions and gives rise to the multiple discrete-continuous (MDC) family of models. These models represent the decision process as a selection of one or more options from a set of alternatives, as well as the decision of the degree of consumption of the chosen alternative(s), subject to linear budget constraints.

The utility function in these models is assumed to be non-linear, quasi-concave, increasing, and continuously differentiable to reflect satiation (i.e., decreasing marginal utility) as consumption increases.

The multiple discrete-continuous extreme value (MDCEV) model, proposed by Bhat (2005), accommodates multiple discreteness based on the generalized variant of the translated constant elasticity of substitution (CES) utility function with a multiplicative log-extreme value distribution for the error term. Moreover, to account for heterogeneity in the population and to produce models that better fit the available data points, population segmentation is proposed in this study. There are two methods for segmentation: exogenous and endogenous. Exogenous segmentation assumes a finite number of mutually exclusive segments, the total number of which is a function of the number of segmentation variables. An apparent setback to this approach is that the number of segments grows dramatically as the number of clustering variables increases. Endogenous segmentation, on the other hand, allows for a large number of segmentation variables to characterize each segment without having the number of segments explode. The parameters on these segmentation variables determine the propensity of belonging to each of the segments and individuals are assigned to segments in a probabilistic manner. Bhat (1997) used the endogenous segmentation approach to segment a population into a finite number of homogenous segments where the utility function is expected to be identical for all individuals probabilistically assigned to a specific segment. However, the utility function is allowed to vary across segments. The number of segments, and the variables that define the segments, are determined as part of the model estimation process. According to Bhat (1997), endogenous segmentation better fits the data as compared to exogenous segmentation, allows for higher order interaction effects, keeps the number of segments under control, and provides more intuitive results with respect to the identification of homogenous clusters of units.

In view of the above, the model used in this thesis is the MDCEV model that accommodates the discrete nature of activity selection as well as the continuous nature of activity participation. To study spatial transferability, the dataset – comprising of states and regions of different socioeconomic composition – is segmented based on a number of spatial characteristics into a number of segments using latent classification. Essentially, regions belonging to the same segment, as a result of latent classification, have a unique model. In other words, parameter equality across regions of the same segment is established.

3.2. Segment-Specific Model Formulation

Assume the dataset is segmented into S homogenous segments where individuals belonging to the same segment s exhibit similar choice behavior, different than those belonging to segment s'. The model considered in this thesis studies activity participation and time-use at the individual-level. All individuals participate in in-home activities and as such, in-home activities are modeled as the outside good in the model structure below – based on a generalized variant of the translated CES utility (Bhat, 2005; Bhat, 2008).

$$U_s(\boldsymbol{x}) = \frac{1}{\alpha_{1s}} \exp(\varepsilon_{1s}) \{(x_1 + \gamma_{1s})^{\alpha_{1s}}\} + \sum_{k=2}^{K} \frac{\gamma_{ks}}{\alpha_{ks}} \psi_{ks} \left\{ \left(\frac{x_k}{\gamma_{ks}} + 1 \right)^{\alpha_{ks}} - 1 \right\} \qquad (1)$$

The first term in this expression corresponds to the utility derived from the consumption of an outside good, i.e., an alternative that is consumed by all individuals in the sample. In its absence, the expression collapses to include just the second term of Equation (1) with k ranging from 1 to K (where k is an alternative). $U_s(\boldsymbol{x})$ is the utility function associated with the consumption quantity x in segment s. It is quasi-concave, increasing, and continuously differentiable with respect to the vector $\boldsymbol{x}$ of dimension (K x 1) ($x_k \geq 0$ for all k alternatives). ψ_{ks} is the baseline marginal utility of consuming good k in segment s, i.e., the utility when there is zero consumption of good k. This utility is expressed in terms of a vector of exogenous variables $\boldsymbol{z_{ks}}$ as follows: $\psi_{ks} = \exp(\boldsymbol{\beta_s'} \boldsymbol{z_{ks}} + \boldsymbol{\varepsilon_{ks}})$ where $\boldsymbol{\beta}$ is a vector of parameters reflecting the sensitivity of the baseline utility to the exogenous variables. The marginal rate of substitution between two goods i and j is the ratio of their baseline marginal utilities. Accordingly, if i and j have the same unit prices, the consumer would gain more utility consuming the alternative with the higher baseline marginal utility and is therefore, more likely to consume that good and prefer it over other goods with similar unit prices.

γ_{ks} is a parameter associated with good k in segment s and plays a dual role. On the one hand, these parameters enable corner solutions (i.e., zero consumption of a good k). On the other hand, these parameters serve as satiation parameters (reflecting preference, analogous to slopes of indifference curves). There is no translation parameter

γ_{1s} associated with the outside good as it is always consumed. α_{ks} is a satiation parameter associated with good *k* in segment *s*. As more of good *k* is consumed, the marginal utility of additional consumption decreases. A value of one for all satiation parameters essentially implies that the consumer does not experience satiation. If there is no satiation effect and if the unit prices of all available goods are the same, the consumer is expected to invest the entirety of his or her budget in the good with the highest baseline marginal utility (i.e., the highest ψ_k value). As the value of α_k decreases from the value of unity, the satiation effect of good *k* increases. The inclusion of both γ_{ks} and α_{ks} in the model specification renders the estimation of Equation (1) impossible as they both reflect satiation behavior. Accordingly, $U_s(\boldsymbol{x})$ can be rewritten in two ways depending the satiation parameter that is estimated (γ_{ks} versus α_{ks}). In the case where the γ_{ks} parameters are estimated as the satiation parameters, $U_s(\boldsymbol{x})$ may be written as:

$$U_s(\boldsymbol{x}) = \exp(\varepsilon_{1s}) \ln\{x_1 + \gamma_{1s}\} + \sum_{k=2}^{K} \gamma_{ks} \exp(\boldsymbol{\beta}_s' \boldsymbol{z}_{ks} + \varepsilon_{ks}) \ln\left(\frac{x_k}{\gamma_{ks}} + 1\right) \quad (2)$$

In the case where the α_{ks} parameters are estimated as the satiation parameters, $U_s(\boldsymbol{x})$ may be written as:

$$U_s(\boldsymbol{x}) = \frac{1}{\alpha_{1s}} \exp(\varepsilon_{1s}) x_1^{\alpha_{1s}} + \sum_{k=2}^{K} \frac{1}{\alpha_{ks}} \exp(\boldsymbol{\beta}_s' \boldsymbol{z}_{ks} + \varepsilon_{ks}) \{(x_k + 1)^{\alpha_{ks}} - 1\} \quad (3)$$

The first terms in equations (2) and (3) refer to the outside good, i.e., in-home activities, in the context of this study. The MDCEV model assumes an extreme value distribution for the error term ε_{ks} and that ε_{ks} is independent of $\boldsymbol{z_{ks}}$ for all goods k. The error terms are also assumed to be independently distributed across alternatives with a scale parameter σ. However, in the absence of information on price variation across the choice alternatives, or when the price is known to be invariant across alternatives, σ can be normalized to one for convenience.

V_{ks} denotes the utility associated with alternative k in segment s and is defined based on two profiles: the γ-profile and the α-profile depicted in Equations (2) and (3) The γ-profile expression of V_{ks} is given as follows:

$$V_{ks} = \boldsymbol{\beta_s'} \boldsymbol{z_{ks}} - \ln\left(\frac{x_{\mathrm{k}}^{*}}{\gamma_{ks}} + 1\right) \tag{4}$$

Equation 4 holds for k = 2, 3, ..., K; $V_{1s} = -\ln\left(x_1^* + \gamma_{1s}\right)$. The α-profile expression of V_{ks} is given as follows:

$$V_{ks} = \boldsymbol{\beta_s'} \boldsymbol{z_{ks}} + (\alpha_{ks} - 1)\ln(x_k^* + 1) \tag{5}$$

Equation 5 holds for k = 2, 3, …, K; $V_{1s} = (\alpha_{1s} - 1)\ln(x_1^*)$. Given the two profiles for the utility V_{ks}, the expression for the probability of the consumption pattern of goods k for individual q (of a total number of individuals Q) conditional on belonging to segment s is as follows:

$$P_q(x_1^*, x_2^*, x_3^*, \dots, x_M^*, 0, 0, \dots, 0)|S$$
$$= \left[\prod_{i=1}^{M} f_i\right]\left[\sum_{i=1}^{M} \frac{1}{f_i}\right]\left[\frac{\prod_{i=1}^{M} e^{V_{is}}}{\left(\sum_{k=1}^{K} e^{V_{ks}}\right)^{M}}\right](M-1)! \tag{6}$$

Where,

$$f_i = \left(\frac{1 - \alpha_i}{x_i^* + \gamma_i}\right) \tag{7}$$

M refers to the total number of consumed goods (M ≥ 1)

x_i^* refers to the consumption quantity of good i.

The individual utility maximization is subject to the budget constraint $\sum_{k=1}^{K} x_k^* = E$ where E is the total continuous quantity available to an individual (24 hours in the context of activity engagement). For convenience, the γ profile is adopted and estimated in this study.

3.3. Segment Assignment Formulation

The latent classification aspect of this model assigns individuals (cities or regions in the context of this model) to the segments. The utility of individual *q* belonging to segment *s* is given by the following expression (Sobhani et al, 2013):

$$W_{qs}^* = \boldsymbol{\delta}_{\boldsymbol{s}}' \boldsymbol{y}_{\boldsymbol{q}} + \xi_{qs} \tag{8}$$

Where,

$\boldsymbol{y_q}$ is a column vector (of dimension *M*x1) of variables, including a constant, that influence the tendency of individual *q* to belong to segment *s*.

$\boldsymbol{\delta_s}$ is a column vector (of dimension *M*x1) of coefficients explaining the sensitivity of the utility W_{qs} to the independent variables $\boldsymbol{y_q}$.

ξ_{qs} is an idiosyncratic random error term assumed to have an independent Type I extreme value distribution across individuals q and segments s.

Accordingly, the probability that individual q belongs to segment s is given as follows:

$$P_{qs} = \frac{\exp(\boldsymbol{\delta}_s' \boldsymbol{y_q})}{\sum_{k=1}^{S} \exp(\boldsymbol{\delta}_k' \boldsymbol{y_q})} \tag{9}$$

Building on Equations 7 and 9, the unconditional probability of the multiple-discrete continuous choice pattern is as follows:

$$P_q = \sum_{s=1}^{S} \left[(P_q(x_1^*, x_2^*, x_3^*, \dots, x_M^*, 0, 0, \dots, 0) | S * P_{qs} \right] \tag{10}$$

Consequently, the likelihood function for the entire dataset (size Q) is as follows:

$$L = \prod_{q=1}^{Q} P_q \tag{11}$$

After determining segment membership, the characteristics of each segment can be obtained by estimating the mean of the variables in each segment as follows (Bhat, 1997):

$$\overline{y_s} = \frac{\sum_q P_{qs} \boldsymbol{y_q}}{\sum_q P_{qs}} \tag{12}$$

3.4. Measures of Goodness-of-Fit

In this study, the model is first estimated assuming the population is comprised of two segments. The number of segments is incrementally increased in a stepwise manner until further segmentation of the population no longer improves goodness-of-fit. The log likelihood value improves as the number of segments increases, calling for the use of more effective goodness-of-fit measures for assessing the optimal number of segments in the dataset. Such measures include the Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Akaike Information Criterion corrected (AICc). The Bayesian Information Criterion (BIC) is given by the following expression (Schwarz, 1978).

$$\text{BIC} = -2LL + K \ln Q \tag{15}$$

where LL is the log likelihood at convergence, K is the number of estimated parameters and Q is the number of observations in the dataset. The Akaike Information Criterion (AIC) is given by the following expression (Akaike, 1974).

$$\text{AIC} = 2K - 2LL \tag{16}$$

The Akaike Information Criterion corrected (AIC_c) is given by the following expression (Sugiura, 1978; Hurvich & Tsai, 1995).

$$\text{AIC}_\text{c} = 2K - 2LL + \frac{2K(K+1)}{Q-K-1} \tag{17}$$

Several studies suggest that the BIC is superior to other assessment measures when it comes to determining the dimensionality of the segment-space (see Rust et al, 1995; Steele and Raftery, 2009). For this reason, the BIC is used in this thesis as the basis for establishing the number of segments S into which the dataset will be divided.

Chapter 4: Model Estimation Results

4.1. Latent Segmentation Results

This section presents the latent segmentation results. A base MDCEV model of activity engagement and time allocation was estimated on the entire data set. In addition, models were estimated assuming a latent segmentation with S=2, 3, and 4 segments. The specifications of each of these models include an array of socio-economic variables (age, gender, household size, income levels, auto ownership) and a contextual variable reflecting area type (urban/rural). The starting values for the two segments model were based on the estimation results of the base MDCEV model (estimated on the entire dataset). The starting values for the three segments model were based on the results of the two segments model and the base MDCEV model. The starting values for the four segments model were based on the results of the three segments model and the base MDCEV model. The BIC was computed for each of the models representing different segmentation schemes as per Equation (15). For the two segments model, the BIC was 326426.7. The value is found to decrease for the three segments model (326049.1), and then increase for the four segments (326053.7). Based on this finding, it may be concluded that three segments is the optimal dimension of the segment-space. It is worth mentioning here that the optimal dimension of the segment-space dimension is dependent on the model and the estimation dataset. In other words, the variation contained within the dataset is what prompted three segments to be the optimal number of segments for this particular model. Future research on different models and datasets may yield a different optimal number of segments. Also note that only the constant satiation parameters were estimated so as not to complicate the model segmentation.

Complete model estimation results for the three-segment MDCEV model are furnished in Table 6, Table 7 and Table 8 in the Appendix. In general, it was found that the model estimation results are intuitive and consistent with expectations. In all segments, an array of socio-economic variables and urban area type influence activity engagement and time allocation. An examination of the signs and magnitudes of some of the variables (gender, area type, number of vehicles in the household) suggests that there is considerable heterogeneity in how individuals of different segments engage their time.

The parameters in the first segment model suggest that the activity with the highest intensity of participation is the active recreation activity. This activity entails going to the gym, exercising, and playing sports. The parameters in the second segment model show that individuals belonging to this segment engage in personal and recreational activities, namely maintenance, social, medical, and other activities, more so than individuals in other segments. The other activity category includes school-related activities, religious activities, relaxation, vacation, family obligations, attending funerals or weddings, pet care, attending meetings, and others. However, overall participation in all of the out-of-home activities in this segment remains less than the participation in in-home activities, suggesting that individuals in this segment are less out-of-home activity oriented when compared with individuals in the other two segments. The parameters in the third segment model indicate that the activity with the highest level of consumption for individuals in this segment is shopping. This includes shopping/running errands and buying goods (e.g., groceries, clothing, and hardware).

Table 2 furnishes estimation results for the latent segmentation portion of the model. This is the model that actually determines the segment into which an individual falls. Once an individual is assigned to a (latent) segment, then the appropriate MDCEV model corresponding to that segment can be used to forecast activity engagement and time use patterns for the specific individual. Residential and employment densities were used as proxies for demographic characteristics as well as area type. Urban versus rural area type is not explicitly introduced into the segmentation configuration due to a high correlation between residential and employment densities on the one hand and urban dummy variable on the other. In addition, a state-specific dummy variable was introduced to account for inter-state versus intra-state transferability. Moreover, transit service quality is represented by the presence or absence of rail in the Metropolitan Statistical Area (MSA) corresponding to the regions in the dataset. Segment 1 is treated as the base in the results furnished in Table 2. A comparison of the parameter signs and magnitudes between the second and third segments provides important qualitative information pertaining to the spatial characteristics of these two segments relative to each other. The model results yield relatively large constants for segments two and three, suggesting that these segments account for a higher share of the sample. Those residing in higher density areas are less likely to fall within segments two and three. However, those residing in high employment density locations are likely to fall within segment two. Individuals in the California data set are less likely to fall in segments two and three, suggesting that there are significant differences between the two states included in this study (Florida and California).

Table 2: The Latent Segmentation Model and Characterization of Three Segments

Segmentation Variable		**Segment 1 (base)**	**Segment 2**	**Segment 3**	**Dataset**
Constants		-	1.6016 (14.56)	1.4904 (13.27)	-
Residential Density (Housing units per sq mi)	< 500 (base)	-	-	-	-
	500 – 1,999	-	-0.2127 (-2.18)	-0.2089 (-2.12)	-
	≥ 2,000	-	-0.1324 (-1.33)	-0.1710 (-1.72)	-
Employment Density (Workers per sq mi)	< 500 (base)	-	-	-	-
	500 – 1,999	-	-	-	-
	≥ 2,000	-	0.1487 (2.70)	-	-
State	California	-	-0.2391 (-3.38)	-0.2709 (-3.74)	-
	Florida (base)	-	-	-	-
Transit Service Quality	Rail (base)	-	-	-	-
	No Rail	-	-0.1450 (-2.18)	-0.1805 (-2.66)	-
Quantitative Characterization of the Three Segments					
Residential Density (Housing units per sq mi)	< 500	14.07%	15.53%	16.31%	15.64%
	500 – 1,999	42.14%	38.99%	40.27%	39.91%
	≥ 2,000	43.79%	45.48%	43.42%	44.45%
Employment Density (Workers per sq mi)	< 500	32.50%	31.84%	33.48%	32.57%
	500 – 1,999	40.89%	38.34%	39.88%	39.28%
	≥ 2,000	26.61%	29.82%	26.64%	28.15%
State	California	70.00%	65.94%	65.16%	66.18%
	Florida	30.00%	34.06%	34.84%	33.82%
Transit Service Quality	Rail	47.46%	50.47%	50.27%	50.00%
	No Rail	52.54%	49.53%	49.73%	50.00%
Area Type	Urban	93.10%	92.60%	92.26%	92.53%
	Rural	6.90%	7.40%	7.74%	7.47%
Share		0.1351	0.4771	0.3878	1.0000

The quantitative characterization of the three segments in Table 2 is performed by computing the mean values of the segmentation variables within each segment as per Equation (12). Overall, it is found that the first segment accounts for about 13.5 percent of the sample, the second segment accounts for 47.7 percent of the sample, and third segment accounts for 38.8 percent of the sample (see the bottom row of Table 2). Within the context of the various characteristics, it is found that segment one is characterized by (individuals living in) areas with higher residential density, low- to medium employment density, and absence of rail service (in comparison to areas that fall into segments two and three). Consistent with this segmentation pattern, the MDCEV model estimation results show that individuals in segment one, who reside in higher residential density neighborhoods as per the segmentation model, are more likely to engage in active recreational pursuits and allocate time to such activities. Similarly, it is found that segment two is largely made up of high density residential and employment areas. The fact that the MDCEV model shows that individuals in this segment engage in a variety of activities such as personal maintenance, social/recreational activities, medical, and other can be attributed to the likelihood that such areas offer diverse and plentiful opportunities for engaging in different kinds of activities. Overall, however, those in segment two pursue out-of-home activities to a lesser degree than those in segments one and three; if they do pursue activities, then it is likely to be a variety of activities as opposed to an emphasis on just one or two activities. It is found that individuals in segment three are likely to fall into lower density areas with presumably fewer opportunities for outdoor pursuits and recreational activities. Consistent with this finding in the latent segmentation model, the MDCEV model shows that individuals in segment three are more prone to undertake shopping activities, presumably because the areas do not offer opportunities for pursuing a variety of different activities.

4.2. Comparison of Endogenous and Exogenous Segmentation Schemes

This section offers a comparison of the performance of the endogenous segmentation scheme versus the traditional exogenous segmentation scheme in which segments are identified based on exogenously defined criteria. It should be noted that the adjusted log likelihood ratio index for the three-segment MDCEV model is 0.4136 and the number of estimated parameters in the model is 325. Table 3 presents results of the comparison showing that the endogenous segmentation scheme outperforms the exogenous segmentation schemes for both one and two-way segmentations.

Traditionally, clusters have been defined by predetermined criteria in order to transfer models within them. Twelve clusters emerge if segments are to be exogenously defined based on one clustering criterion. The expansion of the segmentation dimensionality to two explodes the number of clusters into 57, accounting for all feasible combinations of two-way segmentation. The preferred specification for each of these models was derived by iteratively until significant and behaviorally intuitive parameters remained. The one-way segmentation model results show that area type is the most important segmentation variable with the highest adjusted likelihood ratio index among all one-way segmentation models (0.4094). For this comparison, the adjusted likelihood ratio index for all one-way segmentation models was also computed under the most favorable condition, where the index corresponds to the number of estimated parameters in the base MDCEV model (134 parameters). The resulting $\bar{\rho}^2_{fav}$ values are shown in Table 3 and indicate that, under the most favorable scenario (although unrealistic), the adjusted likelihood ratio index is still less than that of the endogenous segmentation model.

Table 3: Comparison of the Endogenous Segmentation Model with One-Way and Two-Way Exogenous Segmentation Models

Segmentation Variable	Segs	Params	LL at converg	$\bar{\rho}^2$ [2]	$\bar{\rho}^2_{fav}$ [3]	Two-way Segmentation with…	Segs	Params	LL at converg	$\bar{\rho}^2$	$\bar{\rho}^2_{fav}$ [4]
State	2	237	-163160.4833	0.4079	0.4083	Residential Density	6	563	-162805.1273	0.4080	0.4089
						Employment Density	6	597	-147394.6772	0.4079	0.4090
						Transit Service Quality	4	401	-162943.5897	0.4081	0.4084
						Area Type	4	372	-162993.2865	0.4080	0.4082
Residential Density	3	338	-163044.4900	0.4080	0.4087	Employment Density[5]	9	705	-160191.1947	0.4081	0.4095
						Transit Service Quality	6	580	-162790.0652	0.4080	0.4089
						Area Type	6	409	-160412.7488	0.4082	0.4085
Employment Density	3	350	-163032.0739	0.4080	0.4088	Transit Service Quality	6	601	-162768.6610	0.4080	0.4090
						Area Type	6	439	-161115.1576	0.4084	0.4088
Transit Service Quality	2	238	-163159.836	0.4079	0.4083	Area Type	4	304	-159485.0556	0.4082	0.4081
Area Type	2	213	-163165.1216	0.4094	0.4096	-	-	-	-	-	-

[2] The adjusted log likelihood ratio index is computed as $\bar{\rho}^2 = 1 - \frac{\text{LL at convergence} - \text{k}}{\text{LL at zero}}$ where k is the number of estimated parameters.

[3] The favorable adjusted log likelihood ratio index for the one-way segmentation models is computed by replacing k with the number of estimated parameters in the unsegmented model.

[4] The favorable adjusted log likelihood ratio index for the two-way segmentation models is computed by replacing k with the number of estimated parameters in the three-segments model.

[5] The two-way segmentation models between the lowest level of residential density (< 500 housing units per square mile) and the middle and highest level of employment density (500 – 1,999 and ≥ 2,000 workers per square mile respectively), the two-segmentation models between employment density and rural area type, and the two-way segmentation model between the rail transit service quality and the rural area type were not estimable due to small sample size.

Two-way segmentation allows for higher order interaction effects and is expected to better capture preference heterogeneity. In each row, the two-way segmentation corresponds to the pair of variables from the left-most column (one way segmentation variable) and a second variable identified in the middle column. For example, the very first row of the two-way segmentation results correspond to a segmentation based on state and residential density, the second row corresponds to a segmentation based on state and employment density, and so on. The two-way segmentation model results show that the model with the highest adjusted likelihood ratio index is employment density-area type (0.4084). The adjusted likelihood ratio index is then computed for the most favorable scenario where the number of parameters to be substituted in the equation resembles that of the endogenous segmentation model (325 parameters). The resulting $\bar{\rho}^2_{fav}$ values show that the residential-employment density two-way segmentation model outperforms other models (0.4095), but still has a smaller index than that of the endogenous segmentation model (0.4136). Although these values are rather close in magnitude, the efficacy of the two-way exogenous segmentation may be suspect in view of the non-intuitive model parameter estimates obtained in that particular model estimation exercise. For example, some segments showed that people older than 75 years of age engage in fewer medical activities than those belonging to the 65-74 year age group who, in turn, participate in less medical activities than those belonging to the 55-64 year age group, a result that violates a priori expectations. These non-intuitive results can be attributed to the sensitivity of these models to outliers as the small sample sizes within each segmentation scheme amplify the impacts of extreme values.

It is unlikely that only one characteristic of a geographic region deems it similar to another and justifies model transferability between them. Resorting to a higher order segmentation scheme presents issues with the number of segments under consideration as well as the sample size within each segment. In fact, it was not possible to compare the endogenous segmentation model against all possible two-way exogenous segmentation models because of small sample sizes for certain two-way segmentation schemes. This further illustrates the merits of an endogenous segmentation scheme over an exogenous segmentation scheme, the latter being limited by the available sample. Also, the endogenous segmentation scheme offered more behaviorally intuitive model parameter estimates and superior goodness-of-fit, suggesting that it outperforms other exogenous segmentation schemes adopted in prior literature.

4.3. Spatial Transferability Based on Latent Segmentation

Metropolitan areas are different on different levels. The city of Austin may have similar transit service quality as the city of Tampa, but the two are different in residential density. As such, this study proposes that Austin borrow the estimated model as a whole, with all three segments devoted to predict activity generation and time-use, rather than borrowing from an area that reflects the traits of Austin in one or more aspects, but not all. Under this premise, Austin would not have just one model, but a combined model that captures the heterogeneity in the population based on the different criteria identified in the estimated model.

For this purpose, a MDCEV model is estimated for the Austin–San Marcos MSA (sample size 568, cleaned following the same methodology as discussed earlier). The preferred model specification was obtained iteratively by removing insignificant parameters after every model run. The results are furnished in Table 9. The log likelihood of the preferred model is -8655.1. Similarly, the three segments model is applied onto the Austin–San Marcos sample and its log likelihood was found to be -8666.8. The two models are employed to predict the activity consumption levels for the Austin–San Marcos sample. The time predictions using the segmented model were calculated by first determining the probability of each individual belonging to each of the segments and then using these probabilities to weight the activity engagement predictions. Because Equation 2 contains an error term, the calculation of activity engagement for every individual required the generation of a random error term. In order to obtain one activity dwell time for each of the out-of-home activities as well as the in-home activity for every individual, 500 iterations were done in order to average the error out. It is important to note that the error term was generated in the same manner for all three segments in the segmented model. The predicted time-use patterns are found to be quite similar for the two models, with some noted discrepancies for the in-home and shopping activities as shown in Table 4.

Table 4: Comparison of the Predictions by the MDCEV Model for Austin–San Marcos and the Transferred Segmented Model

Activity	**Average (MDCEV – Segmented) (minutes)**
In-home Activity	30.90
Out-of-home Shopping Activity	-42.10
Out-of-home Maintenance Activity	-0.34
Out-of-home Social/Recreational Activity	2.35
Out-of-home Active Recreation Activity	0.28
Out-of-home Medical Activity	0.06
Out-of-home Eating Activity	2.37
Out-of-home Pickup/Drop-off Activity	-0.32
Out-of-home Other Activities	6.79

Figures 1 through 9 in the Appendix show the performance of both the locally estimated model and the transferred model against the corresponding observed activity consumption. The 'MDCEV Model' refers to the model that was estimated using the data from the Austin–San Marcos sample. The 'LC Model' refers to the segmented model that was estimated on the California-Florida dataset and transferred to the Austin–San Marcos MSA. Both models are found to underestimate in-home activity participation. However, on average, the segmented model is found to provide a better approximation for the in-home activity dwell time than the model estimated on local data. It is noteworthy here that the estimation was based on the activity patterns of unemployed adults. This explains the magnitude of in-home activity participation, an issue that is not likely to occur when examining the activity participation of employed adults. Figure 2 shows that the segmented model provides closer out-of-home shopping activity prediction results than the model estimated on the Austin–San Marcos sample. Overall, the figures show that the transferred model performs at least as well as the locally estimated model in all types of activities except for eating out and other activities. The superior performance of the transferred model can be justified by the fact that it caters to heterogeneity in the area make-up, and consequently in activity engagement preferences, by estimating a parameter for every variable per segment. On the other hand, the locally estimated model is comprised of one segment, as activity-travel models typically are, and assigns one parameter to every variable, regardless whether the activities are taking place in a low or high residential density area, or whether the transit service quality in the area endorses activity engagement, and as such, does not quite capture the heterogeneity in the sample and may be more sensitive to the existence of outliers. The transferred model's poor performance in the eating out and other activities predictions can be attributed to a small number of individuals in the California-Florida dataset engaging in these types of activities. This is one limitation of this study as more robust results can be achieved by pooling data from a larger number of sources.

To test the null hypothesis that the transferred model is the true model, a non-nested likelihood ratio test is employed (Koppelman & Bhat, 2006). This test uses the adjusted likelihood ratio index to test that the model with the lower index is the true model. The null hypothesis can be rejected at a significance level given by the following equation:

$$\text{Significance Level} = \Phi\left[-\sqrt{-2\left(\overline{\rho_H}^2 - \overline{\rho_L}^2\right) * \ LL(0) + (K_H - K_L)}\right] \tag{18}$$

Where,

$\overline{\rho_H}^2$ is the adjusted likelihood ratio index for the model with the higher value.

$\overline{\rho_L}^2$ is the adjusted likelihood ratio index for the model with the lower value.

$LL(0)$ is the log likelihood at zero.

K_H is the number of parameters estimated in the model with the higher value.

K_L is the number of parameters estimated in the model with the lower value.

Φ () is the standard normal cumulative distribution function.

The adjusted likelihood ratio index can be calculated as follows.

$$\bar{\rho}^2 = 1 - \frac{\text{LL at convergence} - \text{k}}{\text{LL at zero}} \tag{19}$$

Table 5 shows the adjusted likelihood ratio indexes for the MDCEV model estimated on the Austin–San Marcos sample and for the three-segment MDCEV model that was transferred to the Austin–San Marcos MSA. The table also includes the goodness-of-fit results for models that were estimated for five regions, each similar to the Austin–San Marcos area in only one dimension. The purpose is to show the merits of the new approach over the traditional exogenous transfer approach that is based on identifying an area that is similar to the local region in only one aspect and transferring models between them.

The adjusted likelihood ratio index is calculated for the estimated and transferred models as per Equation 19. The results of the non-nested likelihood ratio test reject the transferred models with a significance level close to zero (Equation 18).

Table 5: Goodness-of-fit Measures for the Locally Estimated and Transferred Models

	A[6]	L[7]	O[8]	J[9]	S[10]	T[11]	W[12]
Number of Parameters	80	325	82	78	76	96	75
Log likelihood at zero	-14752.6						
Log likelihood at constant	-8818.4	-8736.8	-8842.1	-8837.1	-8831.5	-8844.5	-8847.1
Log likelihood at convergence	-8655.1	-8666.8	-8878.5	-8855.5	-8846.8	-8827.7	-8995.6
Rho-squared w.r.t. zero	0.4133	0.4125	0.3982	0.3997	0.4003	0.4016	0.3902
Rho-squared w.r.t. constants	0.0185	0.0080	-	-	-	0.0019	-
Adjusted rho-squared w.r.t. zero	0.4079	0.3905	0.3926	0.3944	0.3952	0.3951	0.3852

The significance level for which the transfer of the three-segments model onto the Austin-San Marcos dataset is rejected can be calculated as follows.

$$\Phi\,[-\,2(0.4079 - 0.3905)\;x\;(-14752.6) + (80 - 325) = \;\Phi\left[-\sqrt{268.39}\right]$$

$$= \;\Phi\,[-\,16.38] \approx 0$$

---

[6] *A* refers to the Austin-San Marcos MDCEV model

[7] *L* refers to the latent class segmented model transferred to Austin-San Marcos

[8] *O* refers to the Orlando MDCEV model transferred to Austin-San Marcos

[9] *J* refers to the Jacksonville MDCEV model transferred to Austin-San Marcos

[10] *S* refers to the Sacramento-Yolo MDCEV model transferred to Austin-San Marcos

[11] *T* refers to the Tampa-St. Petersburg-Clearwater MDCEV model transferred to Austin-San Marcos

[12] *W* refers to the West Palm Beach-Boca Raton MDCEV model transferred to Austin-San Marcos

The 2009 NHTS dataset indicates that both the Austin–San Marcos MSA as well as the Orlando MSA do not have rail. Accordingly, the two are similar in transit service quality. The significance level for which the transfer of the Orlando model onto the Austin–San Marcos dataset is rejected can be calculated as follows.

$$\Phi\,[-2(0.4079 - 0.3926)\;x\;(-14752.6) + (80 - 82) = \Phi\left[-\sqrt{449.43}\right]$$

$$= \Phi\,[-21.20] \approx 0$$

The 2009 NHTS dataset indicates that the Jacksonville MSA has a similar employment density distribution as the Austin–San Marcos MSA. The significance level for which the transfer of the Jacksonville model onto the Austin–San Marcos dataset is rejected can be calculated as follows.

$$\Phi\left[-2(0.4079 - 0.3944)\;x\;(-14752.6) + (80 - 78) = \Phi\left[-\sqrt{400.32}\right]\right.$$

$$= \Phi\,[-20.01] \approx 0$$

The 2009 NHTS dataset indicates that the Sacramento-Yolo MSA is similar to the Austin-San Marcos MSA in terms of percentage of sample falling into the highest employment density category. Accordingly, the two are similar in the employment density dimension. The significance level for which the transfer of the Sacramento-Yolo model onto the Austin–San Marcos dataset is rejected can be calculated as follows.

$$\Phi\,[-2(0.4079 - 0.3952)\;x\;(-14752.6) + (80 - 76) = \Phi\left[-\sqrt{378.72}\right]$$

$$= \Phi\,[-19.46] \approx 0$$

The 2009 NHTS dataset indicates that both the Austin–San Marcos MSA as well as the Tampa-St. Petersburg-Clearwater MSA do not have rail. Accordingly, the two are similar in transit service quality. The significance level for which the transfer of the Tampa-St. Petersburg-Clearwater model onto the Austin–San Marcos dataset is rejected can be calculated as follows.

$$\Phi\,[-\,2(0.4079 - 0.3852)\;x\;(-14752.6) + (80 - 75) = \;\Phi\left[-\sqrt{674.77}\right]$$

$$= \;\Phi\,[-\,25.98] \approx 0$$

The 2009 NHTS dataset indicates that both the Austin–San Marcos MSA as well as the West Palm Beach–Boca Raton MSA do not have rail. Accordingly, the two are similar in transit service quality. The significance level for which transfer of the West Palm Beach-Boca Raton model onto the Austin-San Marcos dataset is rejected can be calculated as follows.

$$\Phi\,[-\,2(0.4079 - 0.3852)\;x\;(-14752.6) + (80 - 75) = \;\Phi\left[-\sqrt{674.77}\right]$$

$$= \;\Phi\,[-\,25.98] \approx 0$$

Although the non-nested likelihood ratio test rejects all models, the significance levels with which the models of the Orlando, Jacksonville, Sacramento-Yolo, Tampa-St. Petersburg-Clearwater and West Palm Beach-Boca Raton MSAs are rejected are larger than that of the segmented model.

The more precise form of the non-nested likelihood ratio test involves the adjusted likelihood ratio indexes with respect to constants (instead of zero). However, it is striking to note that the log likelihoods at constants for the transferred models of the Orlando, Jacksonville, Sacramento-Yolo and West Palm Beach-Boca Raton MSAs are better than those at convergence suggesting that the estimation of parameters to better explain activity engagement patterns is a degradation in the goodness-of-fit of these models. However, this is not the case for the three-segments model.

The non-nested likelihood ratio test is highly sensitive to the inflated value of the log likelihood at zero. Moreover, in absence of data, and given the similar predictive powers of the estimated and transferred models, as shown in Figures 1 through 9, this study suggests that an area, Austin for example, can apply, with caution, the transferred model as is to predict travel behavior.

While it may be better to estimate models based on locally conceived data, borrowing models may be the only resort for planning organizations lacking the resources necessary to undertake data collection. Moreover, as the planning sphere migrates towards activity-based models, the validation of spatial transferability of activity-based models that require large amounts of data and long run times will certainly achieve time and cost savings for planning organizations. Evidently, because the transferred model is segmented to reflect distinct homogenous clusters, it is found to perform well in replicating the predicted times of activity engagement as the model estimated only on data points within the Austin–San Marcos sample.

Chapter 5: Summary and Conclusions

In an era of limited resources and ever-growing demands on disaggregate activity-travel behavior data, metropolitan planning authorities are invariably interested in exploring spatial transferability of models whereby one area may transfer and apply a model estimated in a different, but similar, geographic context.

The traditional approach to identifying an area with a similar profile has been to exogenously (a priori) define a limited set of criteria (say, population and employment size, level and variety of transit service), and then borrow a model from an area that has similar characteristics with respect to the chosen criteria. However, it is difficult to identify the most appropriate set of criteria a priori and the literature has utilized a variety of criteria for transferability, leaving considerable ambiguity for an agency that is seeking to transfer a model from an area with similar population activity-travel characteristics. Rather than approach the transferability paradigm through an exogenous segmentation approach, this study proposes the utilization of an endogenous segmentation approach to help identify similarity measures and create a model that accounts for heterogeneity and can be readily transferred to any local region.

In this thesis, a simultaneous equations model system approach is adopted to accommodate endogenous segmentation. The model system includes a segmentation model coupled with the segment-specific MDCEV model of activity engagement. The latent segmentation model uses a variety of explanatory factors such as area type, transit presence, residential density, and employment density to predict the segment into which an individual would fall, and then the MDCEV model can be used to predict the activity-travel pattern of an individual depending on the segment in which the individual has been placed.

In this study, a NHTS sample including individuals from California and Florida is utilized to estimate the latent segmentation MDCEV model system. It is found that a three-segment model performs best in terms of goodness-of-fit and behavioral intuitiveness. The performance of the endogenous segmentation scheme is found to perform consistently better than alternative exogenous segmentation schemes. The efficacy of the approach is demonstrated through the transfer of the model onto a region that was not part of the estimation dataset, the Austin–San Marcos MSA. The comparison of the predicted time allocations of a locally estimated MDCEV model with that of the transferred model reveals similar predictive powers in replicating observed activity consumptions in the Austin–San Marcos dataset, suggesting that borrowing the segmented model gives reasonable results as compared to estimating models on local data. Moreover, the transferred model outperformed the predictions of the locally estimated model in some instances. The application of a statistical test to study the transferability feasibility rejects that the segmented model can be a true model applied to the Austin–San Marcos MSA. However, the test is found to be dependent on the large log likelihood value at zero, and consequently, rejects the null hypothesis at a significance level close to zero. However, given the good predictive powers of the transferred model to an area with missing or little travel data information justifies the transfer of the model.

The study demonstrates an approach by which similar geographic regions can be identified through an endogenous segmentation process wherein the criteria that define similarity are established within the model estimation phase. This provides a robust mechanism to identify criteria and establish similarity among various regions with respect to the activity-travel characteristics of interest. Future research efforts should be aimed at considering alternative datasets (combining different geographical regions) and different activity-travel characteristics to explore the extent to which the criteria identified in this study vary across datasets and activity-travel dimensions of interest. Moreover, the study remains largely qualitative in nature and future research should be focused on the quantitative and statistical component of the transferability assessment.

Appendix

Table 6: Segment 1 Model Estimation Results

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Baseline Utility Parameters (β)								
Constants	-7.3585 (-20.05)	-9.7802 (-12.62)	-9.2936 (-14.32)	3.4748 (0.98)	-9.8569 (-12.27)	-8.0442 (-19.02)	-9.9352 (-13.94)	-7.0332 (-24.22)
Gender (Female is base)								
Male	-0.3547 (-3.45)	-	-	0.2087 (2.42)	-0.3036 (-1.47)	0.2265 (1.71)	-0.3284 (-1.78)	-
Race (Other is base)								
White	-	0.5161 (2.10)	0.4760 (2.40)	-	-	-	-	0.5304 (2.41)
Black	-	0.4527 (1.11)	-	-0.2772 (-1.28)	-	-0.8569 (-1.95)	0.6588 (1.93)	-
Driver (Non-driver is base)	0.5492 (1.97)	2.1260 (3.13)	0.6128 (1.66)	0.8982 (3.49)	0.8555 (1.34)	0.5607 (1.63)	1.1278 (2.30)	-
Household Size	-0.1058 (-1.50)	-0.1084 (-1.56)	-	-	-	0.1788 (2.52)	0.1536 (1.38)	-
Number of Drivers in Household	0.1834 (1.18)	-	0.3812 (2.02)	0.2673 (2.12)	-	-	-0.3105 (-1.85)	-
Land Use Variable (Rural is base)								
Urban	0.2366 (1.23)	-0.4350 (-2.00)	1.0730 (3.12)	0.2010 (1.29)	-0.5512 (-1.85)	-	-0.3198 (-1.25)	-
Number of Workers	-0.1369 (-1.37)	-	-	-	-	-	0.3355 (2.15)	-
Number of Vehicles in Household	-	-	-	-	-	-	-	-

Table 6 (continued)

	Shop.	**Main.**	**Social/ Rec.**	**Active Rec.**	**Med.**	**Meal**	**Pick/ Drop**	**Other**
Number of Adults	-0.2287 (-1.42)	-	-0.3958 (-2.03)	-0.3796 (-3.04)	-	-0.4265 (-3.32)	-	-
Presence of Children (Youngest child present is 16-21 years is base)								
Youngest child 0-5 years	-	-	-	-	-	-	1.5579 (4.81)	-0.4219 (-1.50)
Youngest child 6-15 years	-	-	-	-	-	-	1.4070 (4.95)	-
Household Income (Less than 25 K is base)								
25 K – 50 K	-	0.2163 (1.20)	-	0.1939 (1.35)	0.8214 (1.69)	-	-	-
50 K – 75 K	0.2508 (1.74)	-	-	0.3461 (2.18)	0.7952 (1.56)	0.3636 (1.93)	-	-
≥75 K	0.2984 (2.50)	0.2837 (1.71)	-	0.3241 (2.25)	1.1533 (2.49)	0.2747 (1.71)	0.4507 (2.63)	-
Education Level (High school level or lower is base)								
Some college level	-	-	-	-	-	0.2479 (1.32)	-	-
Bachelor's level or higher	-	-	-	-	-	0.1886 (1.06)	-	0.4437 (2.84)
Travel Day of the Week (Tuesday-Thursday is base)								
Monday	-0.2896 (-2.19)	-0.3035 (-1.73)	-	-0.1657 (-1.59)	-	-0.6491 (-3.43)	-	-0.5907 (-2.76)
Friday	0.2497 (2.11)	-0.1768 (-1.02)	-	-	-0.3212 (-1.19)	-	-	-0.7988 (-3.45)
Age (Less than 30 is base)								
30 – 54 years	0.5144 (3.42)	-	-0.5269 (-1.67)	-0.6509 (-3.26)	-	-	0.7443 (1.78)	-1.1904 (-4.23)

Table 6 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
55 – 64 years	0.3455 (2.69)	-	-0.3552 (-1.12)	-0.7936 (-3.89)	-	-	0.4987 (1.11)	-1.7168 (-5.60)
65 – 74 years	-	-	-0.3402 (-1.09)	-0.6586 (-3.30)	-	-	0.5798 (1.28)	-1.7869 (-6.21)
≥75 years	-	-	-0.5864 (-1.80)	-0.8069 (-3.84)	-	-0.5009 (-2.87)	-0.8225 (-1.42)	-1.8864 (-6.42)
Satiation Parameters (γ)								
Constants	24.3251 (14.60)	13.1616 (10.08)	100.4211 (9.53)	0.0023 (0.28)	50.0032 (6.70)	35.5211 (10.51)	9.6417 (8.13)	65.4099 (8.15)

Table 7: Segment 2 Model Estimation Results

	Shop.	**Main.**	**Social/ Rec.**	**Active Rec.**	**Med.**	**Meal**	**Pick/ Drop**	**Other**
Baseline Utility Parameters (β)								
Constants	-8.3881 (-29.12)	-8.1817 (-43.11)	-6.9878 (-35.90)	-11.0142 (-25.10)	-8.8377 (-29.72)	-7.4740 (-34.18)	-8.9449 (-45.01)	-5.6693 (-36.96)
Gender (Female is base)								
Male	-0.3967 (-3.94)	-	-0.2026 (-3.22)	0.5511 (3.57)	-0.1261 (-1.71)	-	-0.1363 (-1.66)	-0.1963 (-1.66)
Race (Other is base)								
White	-	-	-	-	-	0.1293 (1.24)	-0.1944 (-2.15)	-
Black	-	-	-	-0.5871 (-1.32)	0.1968 (1.36)	-0.3086 (-1.62)	-	-
Driver (Non-driver is base)	0.7355 (3.05)	0.4887 (3.49)	0.3891 (3.47)	0.5131 (1.27)	-	0.2628 (2.11)	0.7616 (4.62)	-
Household Size	-0.1814 (-3.22)	-	-	-	-0.2536 (-3.16)	-0.1490 (-3.99)	0.1564 (3.08)	-
Number of Drivers in Household	-	-	-	-	-0.2409 (-3.21)	-	-0.2244 (-3.34)	-
Land Use Variable (Rural is base)								
Urban	-	-0.1883 (-1.55)	-	-	0.3401 (2.163)	-0.1771 (-1.54)	-	-0.2478 (-2.14)
Number of Workers	-0.2142 (-2.02)	-0.1851 (-2.98)	-0.1746 (-2.87)	-	-0.1461 (-1.89)	-0.2008 (-3.01)	-	-
Number of Vehicles in Household	-	0.0728 (2.20)	-	-	-	-	-	-
Number of Adults	-	-	-0.1115 (-2.31)	-	0.5184 (4.82)	-	-	-

Table 7 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Presence of Children (Youngest child present is 16-21 years is base)								
Youngest child 0-5 years	-	-0.5791 (-3.63)	-	-	0.3703 (1.79)	-	1.1924 (7.41)	-0.7163 (-5.44)
Youngest child 6-15 years	-	-0.4522 (-3.28)	-	-0.4926 (-1.49)	-	-	1.1331 (8.34)	-
Household Income (Less than 25 K is base)								
25 K – 50 K	-	0.1776 (2.28)	-	0.3601 (1.84)	-0.1716 (-1.80)	0.2631 (2.67)	-	-
50 K – 75 K	-	0.1043 (1.10)	0.2256 (2.70)	-	-0.2290 (-1.94)	0.3986 (3.48)	-	-
≥75 K	-	-	0.2492 (3.51)	0.4503 (2.39)	-0.1433 (-1.31)	0.5977 (5.64)	-	0.1412 (1.98)
Education Level (High school level or lower is base)								
Some college level	0.4960 (3.83)	0.2447 (2.76)	-	0.4634 (2.17)	-	0.1405 (1.65)	0.1382 (1.42)	0.1526 (1.88)
Bachelor’s level or higher	0.2915 (2.21)	0.2053 (2.33)	-	0.7460 (3.57)	-0.1698 (-2.01)	0.1158 (1.32)	0.2123 (2.23)	0.2143 (2.56)
Travel Day of the Week (Tuesday-Thursday is base)								
Monday	-	-	-0.2868 (-3.61)	-	0.1756 (2.02)	-0.1340 (-1.56)	-	-
Friday	0.1686 (1.48)	-	0.1802 (2.39)	-	-0.2610 (-2.54)	0.1098 (1.30)	0.2201 (2.36)	-0.1391 (-1.62)
Age (Less than 30 is base)								
30 – 54 years	0.4787 (3.35)	-	-0.4606 (-3.59)	-	0.7299 (3.28)	-0.3496 (-3.31)	0.4096 (3.77)	-1.5145 (-13.72)
55 – 64 years	0.2697 (1.90)	-	-0.4674 (-3.49)	0.9436 (5.23)	0.9969 (4.46)	-0.3045 (-3.05)	0.3101 (2.68)	-1.8488 (-15.34)

Table 7 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
65 – 74 years	0.3925 (3.19)	-	-0.2060 (-1.60)	0.4423 (2.35)	1.0225 (4.62)	-	-	-1.8710 (-16.73)
≥75 years	-	-	-0.6195 (-4.67)	-	1.0534 (4.76)	-0.2263 (-2.66)	-0.3538 (-2.87)	-2.1357 (-18.97)
Satiation Parameters (γ)								
Constants	25.7513 (15.32)	14.4715 (19.15)	137.2280 (20.71)	1.4452 (9.68)	63.4824 (19.02)	41.2440 (21.87)	14.0651 (16.20)	94.7317 (18.35)

Table 8: Segment 3 Model Estimation Results

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Baseline Utility Parameters (β)								
Constants	3.9347 (1.89)	-7.6414 (-40.58)	-7.9253 (-24.59)	-10.0279 (-58.02)	-8.1778 (-36.33)	-7.8392 (-31.60)	-9.8406 (-29.77)	-7.9331 (-24.26)
Gender (Female is base)								
Male	-0.3350 (-6.76)	-0.1279 (-1.83)	-0.1600 (-1.67)	0.2467 (1.46)	-	-	-0.2342 (-1.91)	-
Race (Other is base)								
White	-0.3346 (-4.75)	-	-0.1498 (-1.28)	-	-	-	-0.2191 (-1.55)	-0.3542 (-2.82)
Black	-0.2284 (-1.79)	0.3829 (2.55)	-	-	-	-	0.2748 (1.10)	-
Driver (Non-driver is base)	-	0.6554 (4.66)	1.0050 (4.41)	-	-	0.2283 (1.38)	0.8417 (3.15)	0.3962 (1.95)
Household Size	-	-	-	-	-	-	0.2880 (4.27)	0.0889 (1.18)
Number of Drivers in Household	-	-0.2260 (-3.84)	-	-	-	-	-0.2128 (-1.73)	-0.2168 (-1.89)
Land Use Variable (Rural is base)								
Urban	-	-0.2630 (-2.18)	-	-	-0.2510 (-1.46)	-	-	-0.2401 (-1.36)
Number of Workers	-	-	0.1553 (1.78)	-	-0.1505 (-1.55)	-0.2237 (-2.84)	0.2640 (2.84)	0.1398 (1.40)
Number of Vehicles in Household	0.0219 (1.02)	0.0764 (2.00)	-	-	-0.1752 (-3.15)	0.0607 (1.55)	-	0.0739 (1.33)
Number of Adults	-	-	-0.1513 (-2.04)	-	0.2203 (2.93)	-	-0.1946 (-1.58)	-

Table 8 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Presence of Children (Youngest child present is 16-21 years is base)								
Youngest child 0-5 years	-	-	-	-	-	-0.3675 (-2.12)	0.6883 (3.10)	-0.2792 (-1.14)
Youngest child 6-15 years	-	-	-	-	-	-	1.3163 (7.07)	-0.2902 (-1.31)
Household Income (Less than 25 K is base)								
25 K – 50 K	-	-	-	-	-	0.1528 (1.16)	0.5445 (2.96)	-
50 K – 75 K	-	-	0.1763 (1.60)	-	-	0.3078 (2.11)	0.6858 (3.40)	-
≥75 K	-	-	-	-	-	0.5049 (3.66)	0.5922 (3.15)	-
Education Level (High school level or lower is base)								
Some college level	-	0.2309 (2.67)	-	-	-	-0.1480 (-1.31)	-	0.2019 (1.58)
Bachelor's level or higher		0.3769 (4.47)	0.3449 (3.75)	0.5849 (3.54)	0.3162 (3.16)	0.1825 (1.68)	0.2298 (1.99)	0.3388 (2.71)
Travel Day of the Week (Tuesday-Thursday is base)								
Monday	-	-	-	-	-	-0.2995 (-2.64)	-0.1905 (-1.44)	-
Friday	-	-	-	-	-0.1895 (-1.52)	-	-	-0.2941 (-2.18)
Age (Less than 30 is base)								
30 – 54 years	-	-	-0.7659 (-4.07)	0.5760 (2.71)	-	-0.3315 (-1.70)	0.3172 (2.31)	-
55 – 64 years	-	-	-0.7472 (-3.89)	-	-	-0.5118 (-2.50)	-	-0.3893 (-2.33)

Table 8 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
65 – 74 years	-	-	-0.8914 (-4.57)	0.5981 (3.14)	-	-0.6738 (-3.28)	-	-0.4077 (-2.53)
≥75 years	-0.1365 (-2.44)	-	-0.8333 (-4.29)	-	-	-0.5898 (-2.84)	-0.5379 (-2.86)	-0.4746 (-2.80)
Satiation Parameters (γ)								
Constants	0.0017 (0.48)	10.9910 (20.67)	94.8695 (13.97)	1.4760 (8.42)	49.3754 (14.50)	39.8343 (16.94)	13.1622 (11.82)	44.4678 (11.51)

Table 9: MDCEV Model Results for Austin–San Marcos

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Baseline Utility Parameters (β)								
Constants	-6.4899 (-26.71)	-7.8784 (-15.46)	-7.2161 (-9.43)	-6.8363 (-15.24)	-7.3508 (-21.21)	-9.1309 (-17.55)	-10.5544 (-12.81)	-4.3147 (-6.80)
Gender (Female is base)								
Male	-0.1872 (-1.33)	-	-	-	-	-	0.3266 (1.34)	-
Race (Other is base)								
White	-	0.7571 (1.74)	-	-	-	1.1295 (2.78)	-	-
Black	-	1.2268 (1.90)	-	-1.3079 (-1.27)	-	-	-	-
Driver (Non-driver is base)	-	-	0.8682 (1.73)	-0.6146 (-1.64)	-	-	1.4476 (1.91)	-
Household Size	-	-	0.1770 (1.29)	-	-	-	0.2646 (1.92)	-
Number of Drivers in Household	-	-0.2700 (-1.80)	-	-0.2504 (-1.51)	-0.6741 (-2.95)	-	-	-
Land Use Variable (Rural is base)								
Urban	-0.1629 (-1.06)	-	0.6355 (2.65)	-0.2657 (-1.25)	-	0.4239 (1.97)	-	-0.4803 (-2.13)
Number of Workers	-	-	-0.4425 (-2.22)	-	-	-0.7179 (-4.17)	0.3634 (1.83)	-
Number of Vehicles in Household	-0.1625 (-2.17)	-	-	-	0.2160 (1.61)	0.1325 (1.39)	-0.4270 (-2.83)	-
Number of Adults	-	-	-0.2267 (-1.09)	-	-	-	-	-0.4358 (-2.58)

Table 9 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
Presence of Children (Youngest child present is 16-21 years is base)								
Youngest child 0-5 years	-0.2992 (-1.25)	-0.7966 (-2.07)	-0.6498 (-1.62)	-	-	-	1.1451 (2.45)	-1.3928 (-3.25)
Youngest child 6-15 years	-	-	-	0.7639 (2.68)	-	-	1.6816 (4.38)	0.4345 (1.37)
Household Income (Less than 25 K is base)								
25 K – 50 K	-	-	-0.4242 (-1.48)	-	-	-	-	0.5431 (2.06)
50 K – 75 K	-	-	-0.3169 (-1.13)	-	-	-	-	0.4923 (2.02)
≥75 K	-	-	-0.4244 (-1.57)	-	-	-	-	-
Education Level (High school level or lower is base)								
Some college level	-	-	-	-	-	0.4002 (1.52)	-	-
Bachelor's level or higher	-	-	-	0.2431 (1.23)	-	0.3819 (1.56)	0.3923 (1.73)	-
Travel Day of the Week (Tuesday-Thursday is base)								
Monday	-	-	-0.5859 (-2.24)	-	-	-0.5118 (-2.03)	-	-0.7021 (-2.41)
Friday	0.2528 (1.61)	0.3025 (1.46)	-	-	-	0.6203 (3.10)	-	-
Age (Less than 30 is base)								
30 – 54 years	0.6024 (2.94)	-	-1.0262 (2.68)	-	-	-	-	-2.1042 (-5.43)
55 – 64 years	0.5604 (2.75)	-	-1.0916 (-2.69)	-	-	-	0.6322 (2.13)	-2.9941 (-6.76)

Table 9 (continued)

	Shop.	Main.	Social/ Rec.	Active Rec.	Med.	Meal	Pick/ Drop	Other
65 – 74 years	0.4954 (2.64)	-	-1.3844 (-3.41)	-	-	-	-	-2.8206 (-6.79)
≥75 years	-	-	-1.1886 (-2.84)	-	-	-0.7797 (-3.02)	-	-3.1875 (-6.96)
Satiation Parameters (γ)								
Constants	24.9321 (10.34)	10.9674 (7.18)	106.0662 (6.67)	36.0612 (5.74)	62.0493 (5.84)	41.7506 (7.57)	10.5775 (5.69)	84.2684 (5.77)

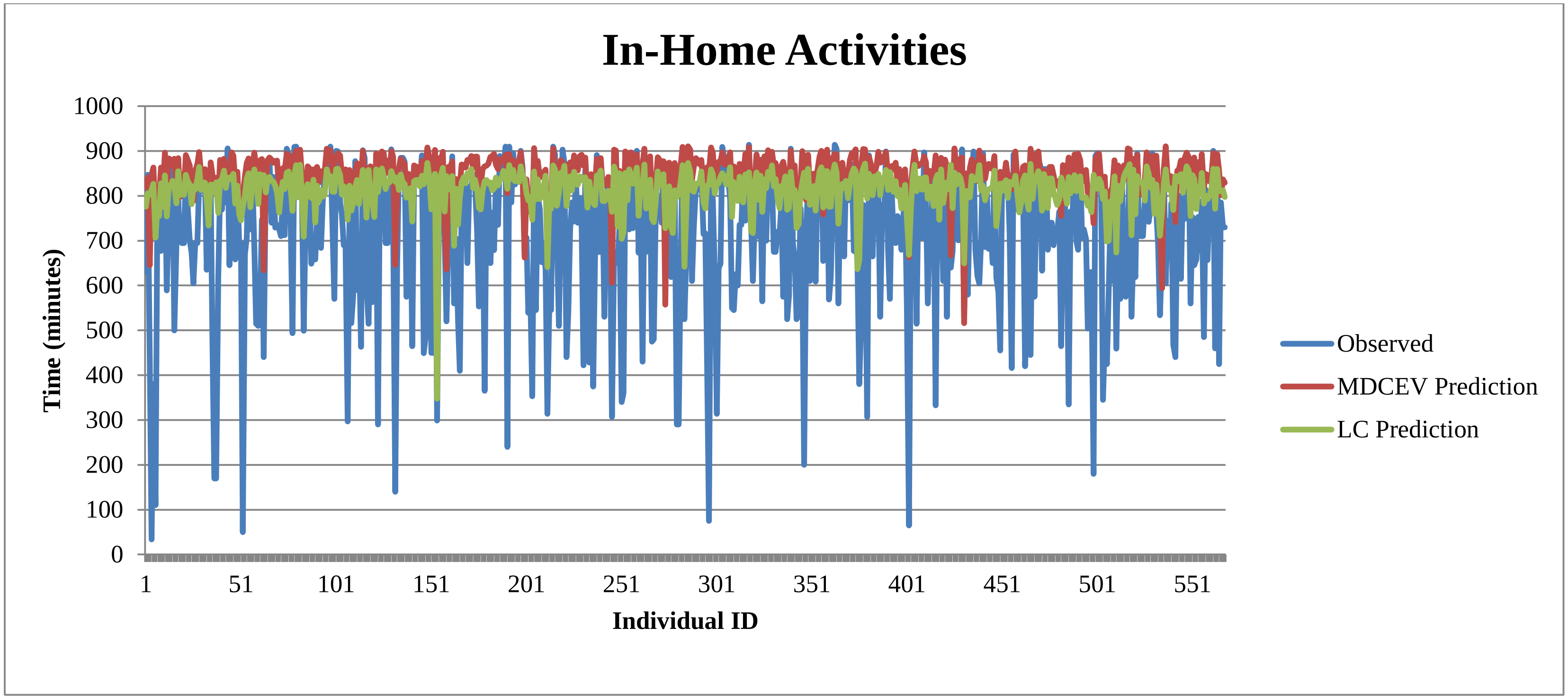

Figure 1: Comparison of Predictions of In-Home Activities by the Estimated and Transferred Model against the Observed In-Home Activity Participation

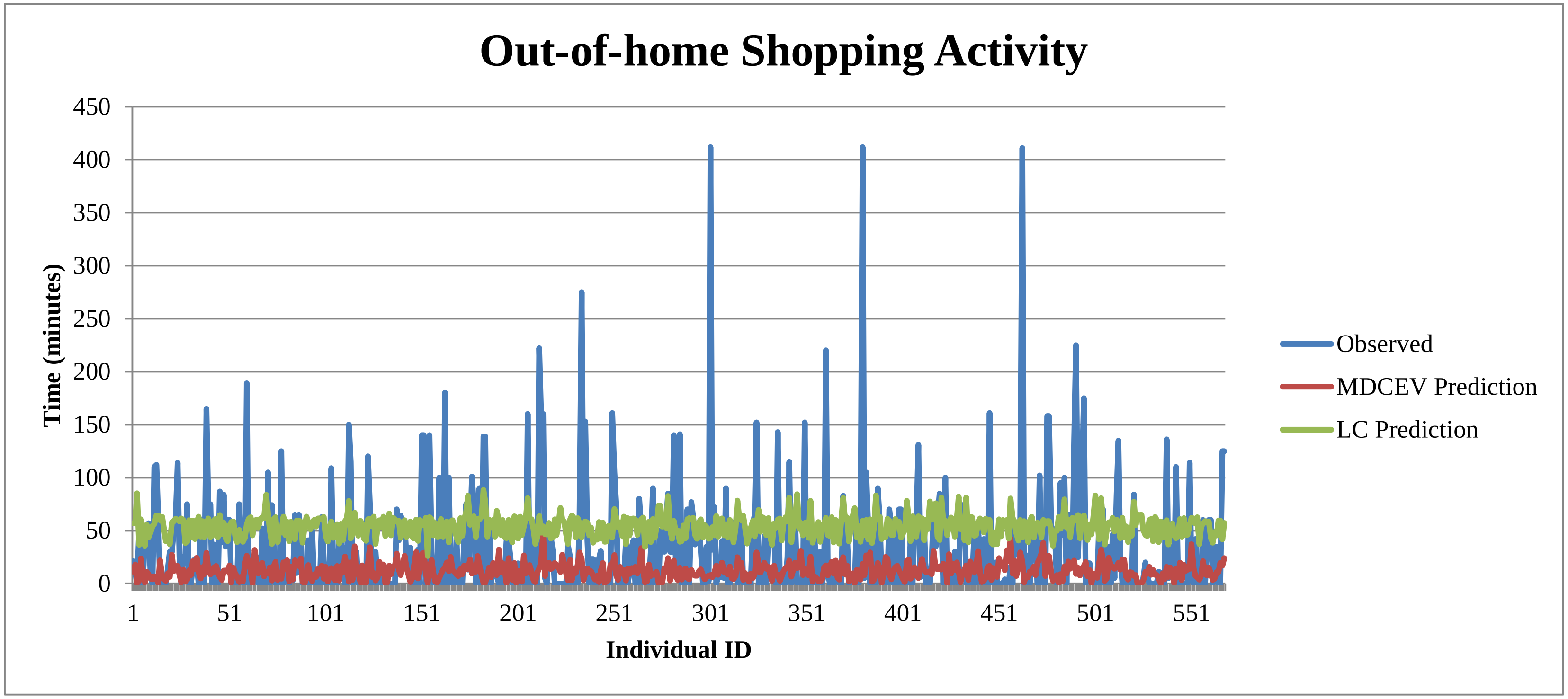

Figure 2: Comparison of Predictions of Out-of-Home Shopping Activities by the Estimated and Transferred Model against the Observed Out-of-Home Shopping Activity Participation

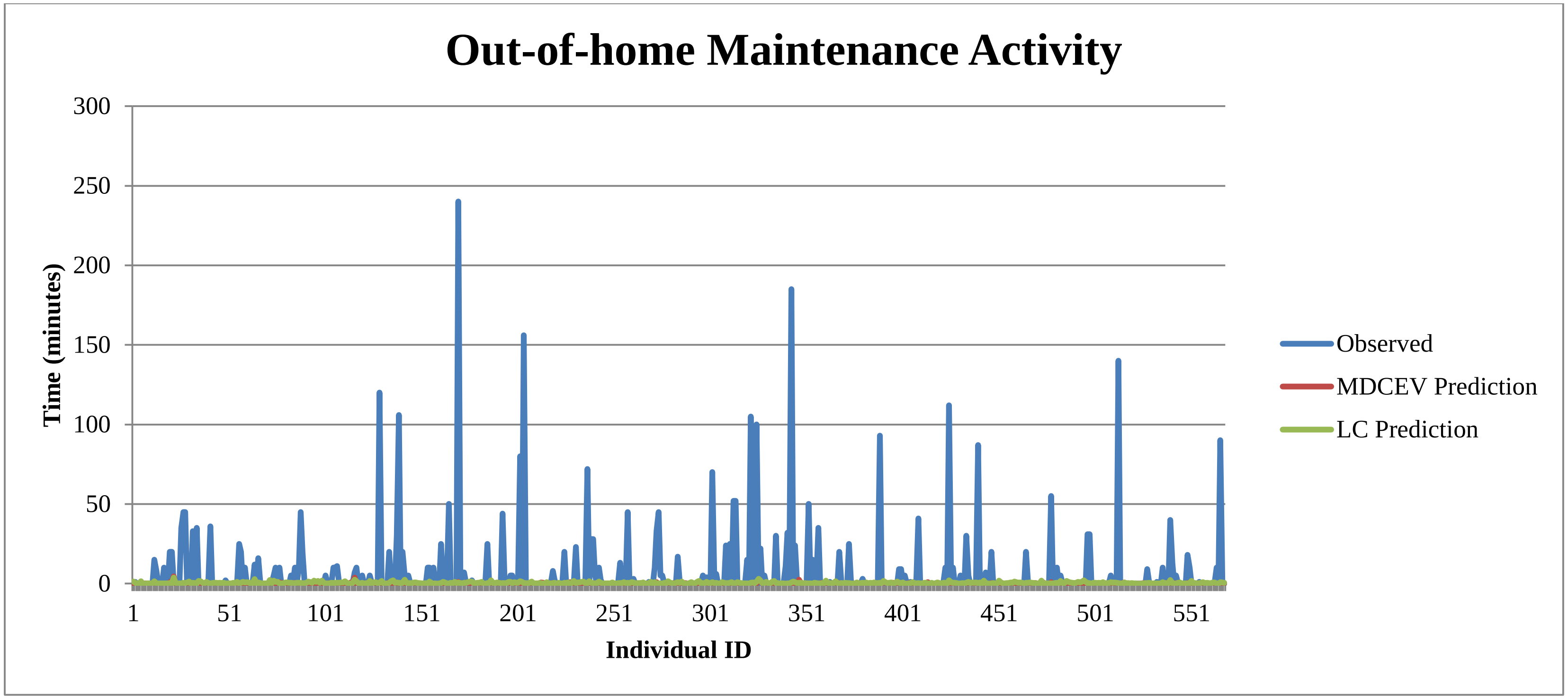

Figure 3: Comparison of Predictions of Out-of-Home Maintenance Activities by the Estimated and Transferred Model against the Observed Out-of-Home Maintenance Activity Participation

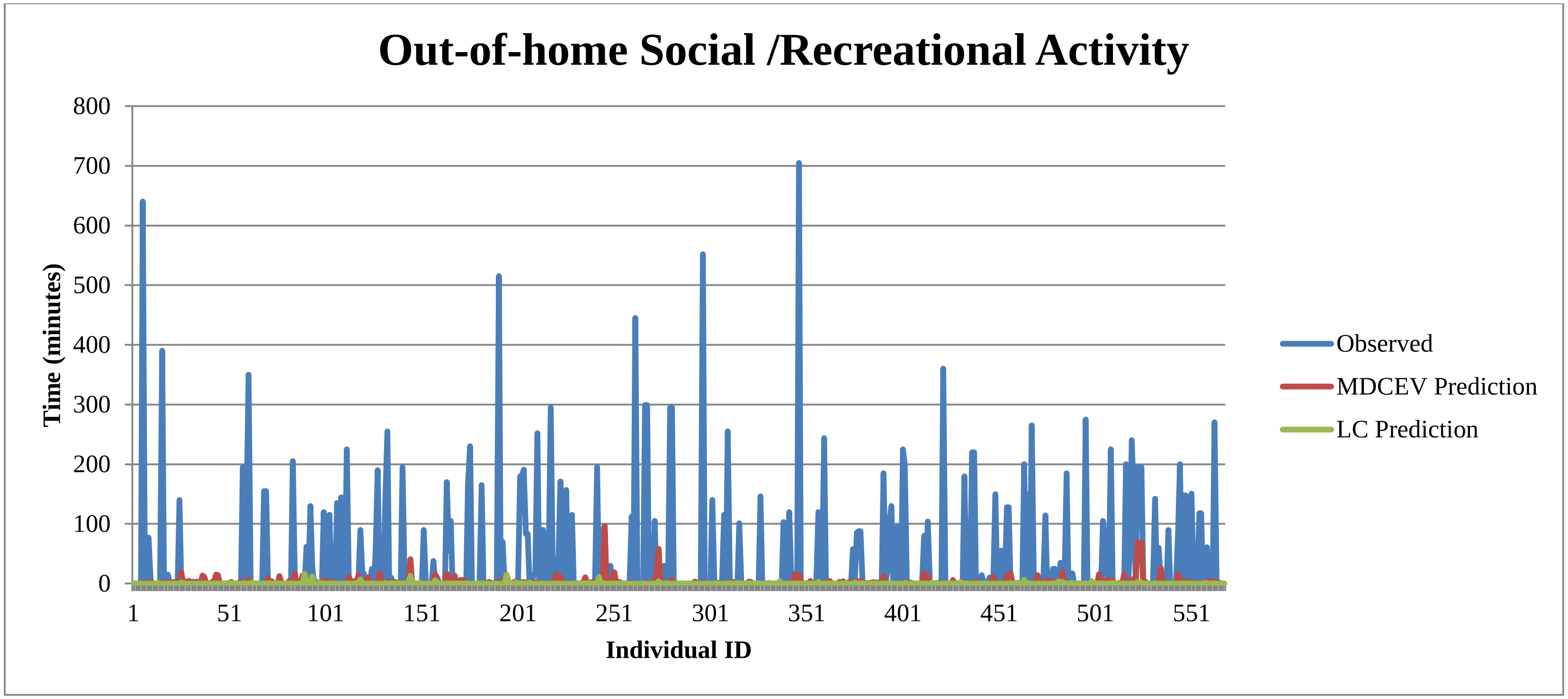

Figure 4: Comparison of Predictions of Out-of-Home Social/Recreational Activities by the Estimated and Transferred Model against the Observed Out-of-Home Social/Recreational Activity Participation

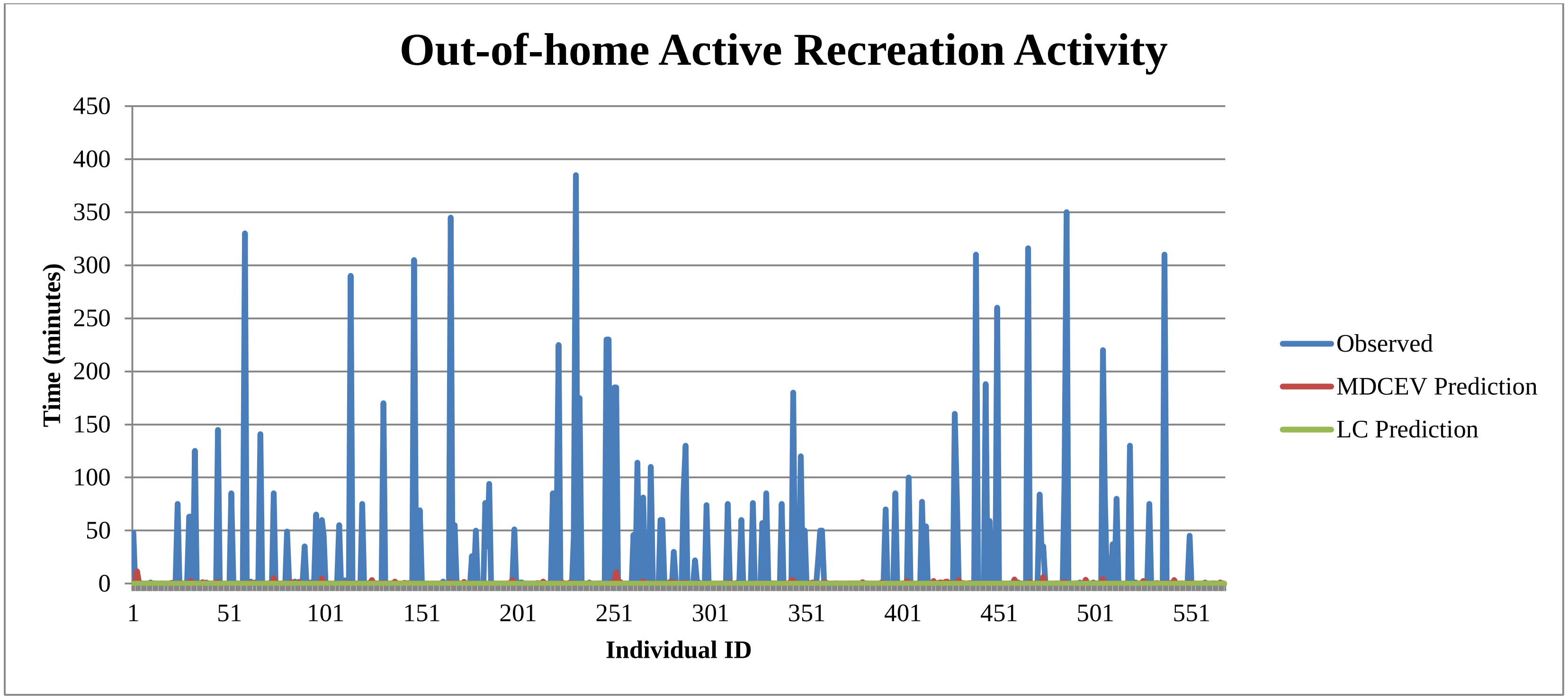

Figure 5: Comparison of Predictions of Out-of-Home Active Recreation Activities by the Estimated and Transferred Model against the Observed Out-of-Home Active Recreation Activity Participation

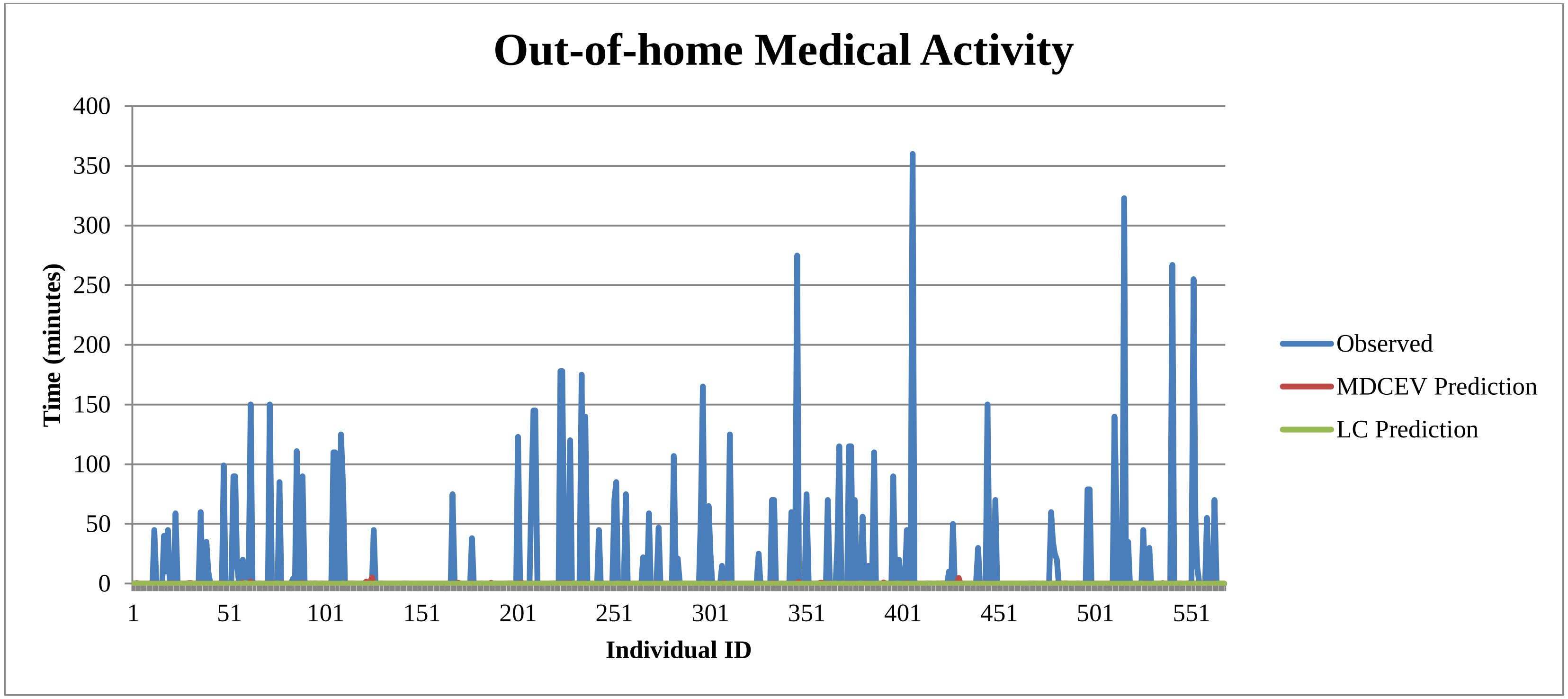

Figure 6: Comparison of Predictions of Out-of-Home Medical Activities by the Estimated and Transferred Model against the Observed Out-of-Home Medical Activity Participation

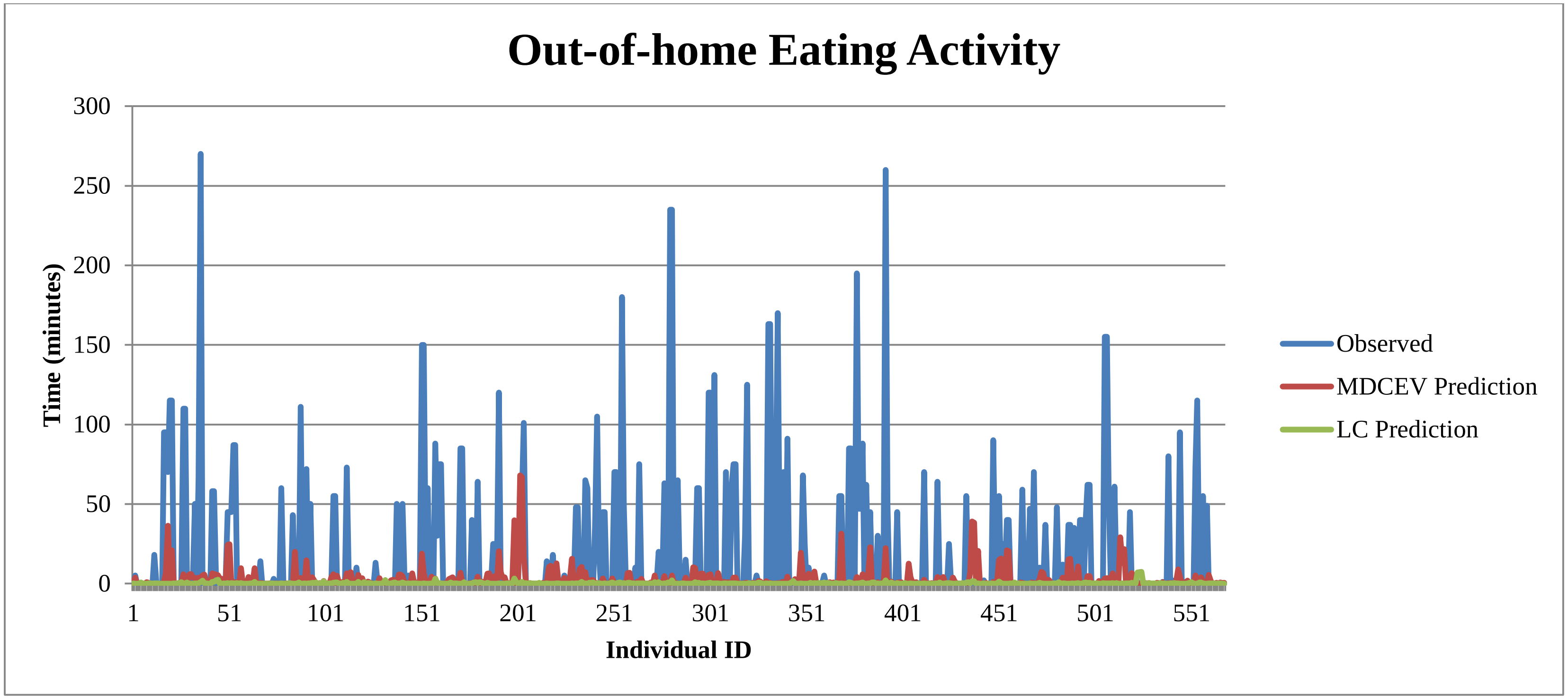

Figure 7: Comparison of Predictions of Out-of-Home Eating Activities by the Estimated and Transferred Model against the Observed Out-of-Home Eating Activity Participation

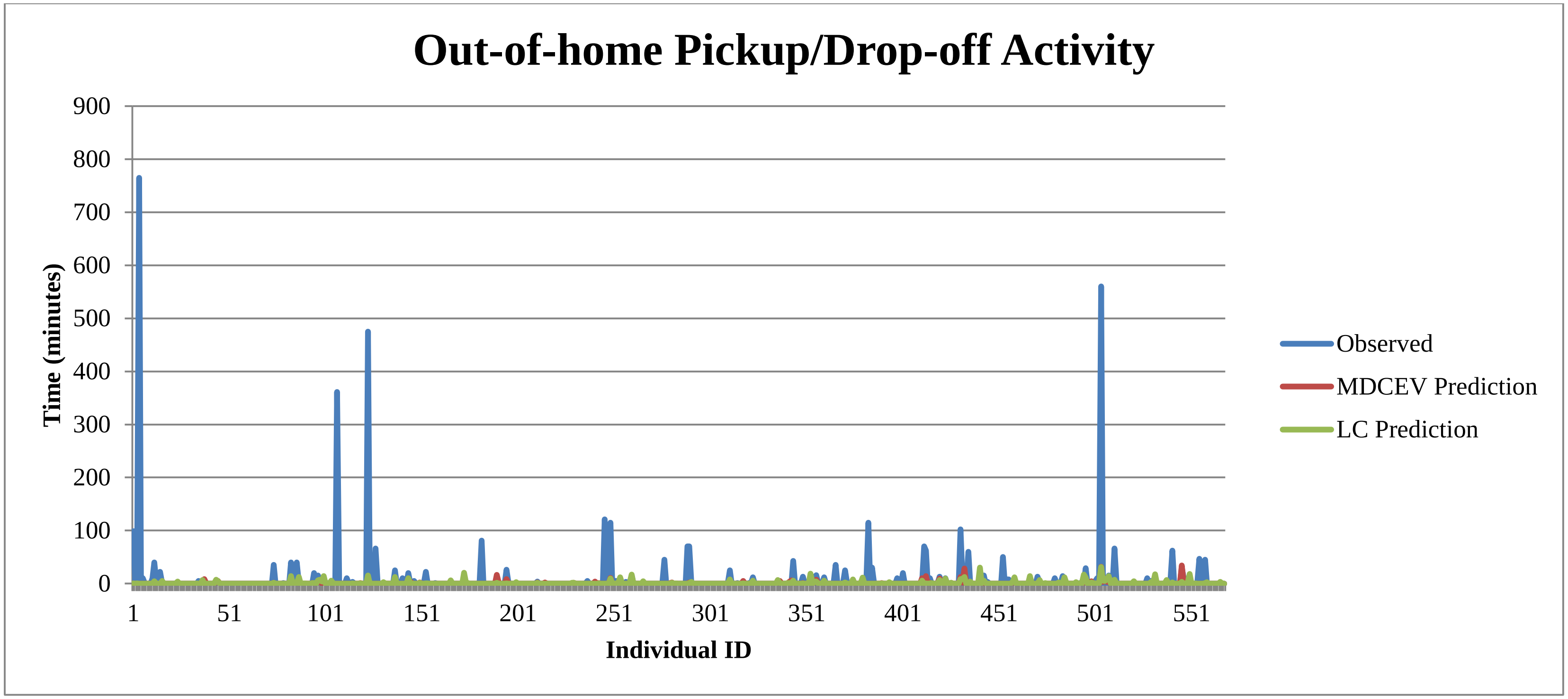

Figure 8: Comparison of Predictions of Out-of-Home Pickup/Drop-off Activities by the Estimated and Transferred Model against the Observed Out-of-Home Pickup/Drop-off Activity Participation

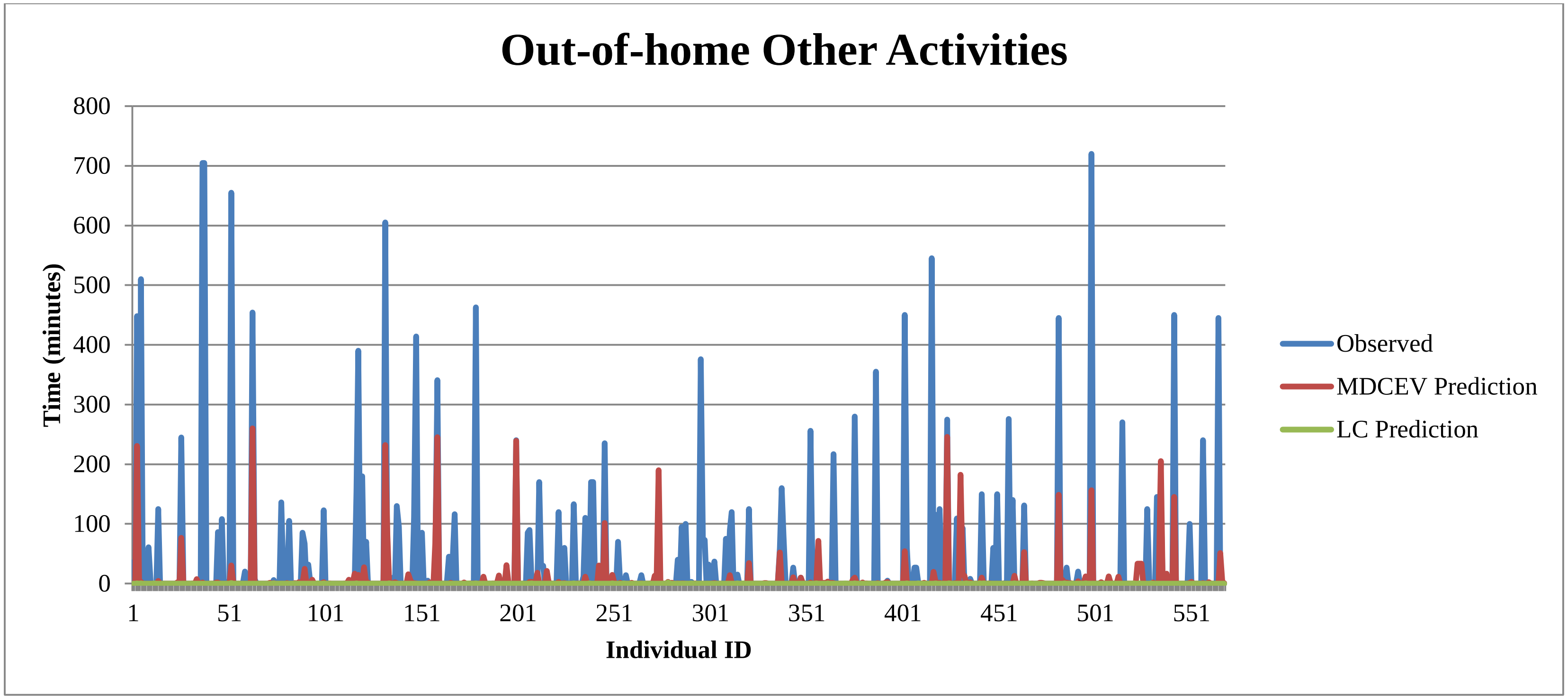

Figure 9: Comparison of Predictions of Out-of-Home Other Activities by the Estimated and Transferred Model against the Observed Out-of-Home Other Activity Participation

Bibliography

1. Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, 19*(6).
2. Arentze, T., Hofman, F., van Mourik, H., & Timmermans, H. (2002). Spatial Transferability of the Albatross Model System: Empirical Evidence from Two Case Studies. *Transportation Research Record 1805*, 1-7.
3. Bekhor, S., & Prato, C. G. (2009). Methodological Transferability in Route Choice Modelling. *Transportation Research Part B, 43*, 422-437.
4. Bhat, C. R. (1997). An Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science, 31*(1), 34-48.
5. Bhat, C. R. (2005). A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B, 39*(8), 679-707.
6. Bhat, C. R. (2008). The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research Part B, 42*(3), 274-303.
7. Bowman, J. L., Bradley, M., Castiglione, J., & Yoder, S. L. (2014). Making Advanced Travel Forecasting Models Affordable Through Model Transferability. *Transportation Research Board 93rd Annual Meeting*. Washington, D.C.: Transportation Research Board.

8. Caldwell, L. C., & Demetsky, M. J. (1980). Transferability of Trip Generation Models. *Transportation Research Record 751*, 56-62.
9. Everett, J. D. (2009). *An Investigation of the Transferability of Trip Generation Models and the Utilization of a Spatial Context Variable.* Knoxville: University of Tennessee.
10. Hurvich, C. M., & Tsai, C.-L. (1995). Model Selection for Extended Quasi-Likelihood Models in Small Samples. *Biometrics*, 1077-1084.
11. Karasmaa, N. (2001). The Spatial Transferability of the Helsinki Metropolitan Area Mode Choice Models. *presented at the 5th workshop of the Nordic research network on modeling transport, land-use, and the environment*, (pp. 1-24).
12. Koppelman, F. S., & Bhat, C. (2006). *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.* U.S. Department of Transportation Federal Transit Administration.
13. Koppelman, F. S., & Pas, E. I. (1986). Multidimensional Choice Model Transferability. *Transportation Research Part B, 20B*(4), 321-330.
14. Koppelman, F. S., & Rose, J. (1983). Geographic Transfer of Travel Choice Models: Evaluations and Procedures. *Optimization and Discrete Choice in Urban Systems*, (pp. 272-309).
15. Koppelman, F. S., & Wilmot, C. G. (1982). Transferability Analysis of Disaggregate Choice Models. *Transportation Research Record*, *895*, pp. 18-24. Washington, D.C.

16. Koppelman, F. S., & Wilmot, C. G. (1986). The Effect of Omission Variables on Choice Model Transferability. *Transportation Research Part B, 20B*(3), 205-213.

17. McComb, L. A. (1986). Analysis of the Transferability of Sisaggregate Demand Models Among Ten Canadian Cities. *Tribune Des Transports*, *3*(1), 19-32.

18. Mohammadian, A., & Zhang, Y. (2007). Investigating the Transferability of National Household Travel Survey Data. *Transportation Research Record 1993*, 67-79.

19. Nowrouzian, R., & Srinivasan, S. (2012). Empirical Analysis of Spatial Transferability of Tour-Generation Models. *Transportation Research Record 2302*, 14-22.

20. Reuscher, T. R., Schmoyer, R. L., & Hu, P. S. (2002). Transferability of Nationwide Personal Transportation Survey Data to Regional and Local Scales. *Transportation Research Record 1817*, 25-35.

21. Rust, R., Simester, D., Brodie, R., & Nilikant, V. (1995). Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combinations of Criteria. Management Science, 41(2), 322-333.

22. Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.

23. Sikder, S., & Pinjari, A. R. (2013a). Spatial Transferability of Person-Level Daily Activity Generation and Time-Use Models: An Empirical Assessment. *Transportation Research Record 2343*, 95-104.

24. Sikder, S., Augustin, B., Pinjari, A. R., & Eluru, N. (2013b). Spatial Transferability of Tour-based Time-of-day Choice Models: An Empirical Assessment. *Procedia - Social and Behavioral Sciences, 104*, 640-649.

25. Sikder, S., Pinjari, A. R., Srinivasan, S., & Nowrouzian, R. (2012). Spatial Transferability of Travel Forecasting Models: A Review and Synthesis. *4th Innovations in Travel Modeling Conference.*

26. Sobhani, A., Eluru, N., & Faghig-Imani, A. (2013). A Latent Segmentation Based Multiple Discrete Continuous Extreme Value Model. *Transportation Research Part B, 58*, 154-169.

27. Steele, R., & Raftery, A. (2009). Performance of Bayesian Model Selection Criteria for Gaussian Mixture Models. Seattle: University of Washington.

28. Stopher, P. R., Greaves, S., & Bullock, P. (2003). Simulating Household Travel Survey Data: Application to Two Urban Areas. *Transportation Research Board 82nd Annual Meeting.* Washington, D.C.: Transportation Research Board.

29. Sugiura, N. (1978). Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections. *Communications in Statistics, Theory and Methods, 7*, 13-26.

30. Wilmot, C. (1995). Evidence of Transferability of Trip Generation Models. *Journal of Transportation Engineering, 121*(5), 405-410.